DOCUMENT RESUME

ED 039 893 LI 001 946

AUTHOR	Seiden, Herbert R.
TITLE	A Comparative Analysis of Interactive Storage and Retrieval Systems with Implications for BCN Design.
INSTITUTION	System Development Corp., Santa Monica, Calif.
SPONS AGENCY	National Library of Medicine, Bethesda, Md.
REPORT NO	TM-4421
PUB DATE	12 Jan 70
NOTE	65p.
EDRS PRICE	EDRS Price MF-$0.50 HC-$3.35
DESCRIPTORS	*Automation, *Comparative Analysis, *Information Networks, *Information Retrieval, *Information Storage, Medical Libraries, Systems Development
IDENTIFIERS	BCN, *Biomedical Communications Networks, National Library of Medicine, NLM

ABSTRACT

The work reported is part of a series of studies aimed at providing information and assistance to the National Library of Medicine (NLM) in planning the Biomedical Communications Network (BCN). The first part reviews the literature on systems relevant to BCN design, documents System Development Corporation's position with respect to certain concepts of bibliographic retrieval as they relate to BCN planning and provides a basis for better understanding of the comparisons. The second part contains comparisons of various systems based on data available in open sources. Over 150 systems were reviewed. Of these, 26 general-purpose and 11 bibliographic systems were selected for comparison, based on criteria described in the study. Implications for system design, for the BCN user, and for network planning are discussed in the third part. Some of the major points are: (1) NLM should stay with the bibliographic retrieval design; (2) a unified network concept should be developed and implemented; (3) there is a need for standards and specifications for inputs, thesauri and unit records: (4) users should be provided more direct access to the files on an interactive basis and (5) some files should be maintained centrally while others should be duplicated at several centers. (Author/NH)

EDO 39893

TM-4421

TECHNICAL MEMORANDUM

(TM Series)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED 039946

A COMPARATIVE ANALYSIS OF INTERACTIVE STORAGE AND RETRIEVAL SYSTEMS WITH IMPLICATIONS FOR BCN DESIGN

Herbert R. Seiden

January 12, 1970

SYSTEM DEVELOPMENT CORPORATION
2500 COLORADO AVE.
SANTA MONICA
CALIFORNIA
90406

This document was produced under Contract NIH-69-71, entitled Design Support for On-Line Retrieval in the Biomedical Communications Network, for the National Library of Medicine, U.S. Department of Health, Education, and Welfare.

PREFACE

This report is submitted to the National Library of Medicine as a requirement of contract 69-71. The study reviews interactive information storage and retrieval systems and presents implications from those reviews for BCN development. I did not attempt to review all aspects of all systems (there are several commercially available documents that do so comprehensively--they are cited in the Bibliography). Rather, my focus was upon those aspects and those systems which are significant for the BCN.

I wish to especially thank Donald Blankenship and Robert V. Katter for their invaluable assistance.

Herbert R. Seiden

ABSTRACT

This document presents a review and analysis of bibliographic retrieval concepts, comparisons of bibliographic retrieval systems, and implications of the review and the comparisons for the design, development, and implementation of the Biomedical Communications Network (BCN). The work reported on here is part of a series of studies aimed at providing information and assistance to the National Library of Medicine in planning the BCN.

The first part of this report reviews the available literature on systems relevant to BCN design, documents SDC's position with respect to certain concepts of bibliographic retrieval as they relate to BCN planning, and provides a basis for better understanding the comparisons. System inputs and file characteristics are shown to be related to system design philosophy. As opposed to batch-mode search, on-line interactive search conducted directly by the ultimate user allows him to quickly reformulate his search in terms of the realities of the data base. To exploit this concept fully, some specific capabilities are required. These include a rich set of search strategies; free use of Booleans; the ability to enter commands, qualify retrievals, and link search statements; the ability to browse or explore related index terms; and provision for relatively unrestricted access to the files.

The second part of this report contains comparisons of various systems, based primarily on data available in open sources. Over 150 systems were initially reviewed. Of these, 26 general-purpose and 11 bibliographic systems were selected for comparison, based upon criteria described in the study. The comparisons indicate the types of systems available for bibliographic retrieval and the capabilities of such systems. They also provide a source of system design criteria for the BCN. The comparisons of the general-purpose systems highlight their basic similarities and suggest that they are too limited, with respect to the type of services required in the BCN. Two patterns or types of bibliographic retrieval systems emerge from the comparison of such systems--one oriented toward bibliographic retrieval, the other toward full-text. The study suggests that the former is more appropriate for BCN purposes.

Implications for system design, for the BCN user, and for network planning are discussed in the third part. Some of the major points are: that NLM should stay with the present design of bibliographic (as opposed to full-text) retrieval; that a unified network concept should be developed and implemented (as opposed to linking existing systems together); that there is a need for standards and specifications for inputs, thesauri, and unit records; that users should be provided more direct access to the files, on an interactive basis, than is presently planned; and that some files should be maintained centrally while others should be duplicated at several centers.

LIST OF FIGURES AND TABLES

Page

FIGURE 1. CONCEPTUAL MODEL OF A BIBLIOGRAPHIC STORAGE AND RETRIEVAL SYSTEM. 3

FIGURE 2. MATRIX FOR RELATING CRITICAL FACTORS, UNIT RECORD CONTENTS, AND OPERATIONAL MODE. 6

FIGURE 3. RELATION OF FILES FOR PROCESSING SEARCH STATEMENTS 8

FIGURE 4. HYPOTHETICAL DISTRIBUTION CURVES INDICATING NUMBER OF TITLE WORDS AND AUTHOR NAMES AS A FUNCTION OF NUMBER OF DOCUMENTS 11

TABLE 1. FILE SIZES AND UNIT RECORD CAPACITY FOR A TOTAL FILE CAPACITY OF 300 X 10^8 BYTES. 12

TABLE 2. LIST AND BRIEF DESCRIPTION OF SELECTED GENERAL-PURPOSE SYSTEMS 25

TABLE 3. OVERVIEW OF SELECTED BIBLIOGRAPHIC RETRIEVAL SYSTEMS 40

TABLE 4. TECHNICAL FEATURES. 46

TABLE 5. SEARCH STRATEGIES 48

TABLE 6. ON-LINE OPERATING CHARACTERISTICS 50

TABLE 7. ERROR CONTROL CAPABILITIES. 52

C O N T E N T S

		Page
1.	INTRODUCTION	1
2.	BIBLIOGRAPHIC RETRIEVAL: REVIEW AND ANALYSIS	2
2.1	A GENERALIZED SYSTEM MODEL	2
2.2	INDEXING AND SYSTEM INPUTS	4
2.3	SELECTED COMPUTER AND SOFTWARE CHARACTERISTICS	5
2.3.1	FILE CHARACTERISTICS	5
2.3.2	OPERATIONAL CHARACTERISTICS	14
2.3.3	PERFORMANCE MEASURES	15
2.4	USER CONSIDERATIONS	16
2.5	NETWORK FACTORS	19
3.	SYSTEM COMPARISONS	21
3.1	GENERAL-PURPOSE DATA MANAGEMENT SYSTEMS	22
3.2	BIBLIOGRAPHIC DATA RETRIEVAL SYSTEMS	35
3.2.1	SELECTION CRITERIA	35
3.2.2	OVERVIEW	37
3.2.3	DETAILED COMPARISONS	38
3.3	COMMENTS ON THE COMPARISONS	54
4.	IMPLICATIONS FOR THE BCN	56
4.1	IMPLICATIONS FOR SYSTEM DESIGN	56
4.2	IMPLICATIONS FOR THE BCN USER	57
4.3	IMPLICATIONS FOR NETWORK PLANNING	57
5.	BIBLIOGRAPHY	59

A COMPARATIVE ANALYSIS OF INTERACTIVE STORAGE AND RETRIEVAL SYSTEMS WITH IMPLICATIONS FOR THE BCN

1. INTRODUCTION

This report compares the characteristics of on-line, interactive information storage and retrieval systems, with particular emphasis on bibliographic applications. The comparisons, which are based primarily on information compiled from open document sources (see Bibliography), provide one basis for determining those features or techniques that should be considered in the design and planning for the Biomedical Communications Network (BCN).

There are three main sections in this report: Section 2, Bibliographic Retrieval: Review and Analysis; Section 3, Systems Comparisons; and Section 4, Implications for the BCN. The review and analysis in Section 2 describes the many elements of bibliographic retrieval systems that affect performance of such systems. The section starts with a generalized model of information storage and retrieval systems describing three major aspects of such systems: input, central processor, and system user. The model is used to explain basic concepts of indexing and system inputs, computer and software characteristics, operational characteristics, user considerations in system design, network factors to consider in BCN planning, and the relationship among these factors. Such details as the relationship of file structure and size to search strategies are included. This provides a basis for better understanding the systems comparisons which follow. In Section 3, a number of systems that are relevant to BCN design are compared. Two major categories of systems are included: general-purpose data management systems, and bibliographic storage and retrieval systems. The latter category is of special interest for BCN planning. Included in this section are the criteria that were used to determine which candidate systems to include in this report, the overall features of those systems, and comparisons of the systems in terms of their implications potential for or use in the BCN. All the cases presented represent operating or near-operational systems. Thus, the feasibility of the features or techniques reviewed is not at issue. Section 4 presents conclusions derived during this study and indicates implications for BCN development arising from those conclusions. In particular, some guidelines for BCN planning are presented.

2. BIBLIOGRAPHIC RETRIEVAL: REVIEW AND ANALYSIS

In order to provide a basis for understanding the features of the systems compared in Section 3, we present a review and analysis of selected features of bibliographic retrieval systems. The emphasis is on those features that are particularly relevant for BCN planning and design. The analysis also serves as a basis for developing guidelines necessary for evaluating individual systems in terms of their compatibility with a central network plan.

This section contains five major parts: A Generalized System Model, Indexing and System Inputs, Computer and Software Characteristics, User Considerations, and Network Factors. The model reviews some basics and serves as a foundation for subsequent discussions. The next three parts analyze, in considerable detail, those bibliographic retrieval system characteristics considered important for BCN planning. The part on networks describes some problems with respect to networking bibliographic retrieval systems.

2.1 A GENERALIZED SYSTEM MODEL

Figure 1 shows the basic organization of a typical bibliographic storage and retrieval system. To facilitate discussion, in Figure 1 the storage and retrieval system model is divided into three functional elements: input, central processor, and system use. As used here, "input" refers to those processes that are performed in order to create the files which store the information for later retrieval. The major functions here are indexing the documents and creating unit records for them. "Central processor" refers to the computer, its peripheral equipment, and the software. The major functions are processing user requests, searching or otherwise manipulating the files, and generating outputs. "System Use" refers to the person(s) using the system and the associated user equipment, procedures, and languages. The major function here is the user/system interface, i.e., inputting requests and commands and displaying outputs. Other user/system interfaces may exist at the input end of the system (where a user might modify or introduce inputs) and at the central processor (where the user might obtain off-line outputs).

The computer programs operate under some "operating system" or "executive program" environment. From a user viewpoint the operating system or executive is of little or no concern; for network planning it becomes very important. Therefore, it is discussed in Section 2.2, Network Factors.

While the model in Figure 1 is sufficiently general to encompass almost any information storage and retrieval system, the discussion throughout the remainder of Section 2 will focus on bibliographic retrieval systems. Their relation to other systems, such as general-purpose file-management systems and library technical processing systems, is discussed in Section 3.

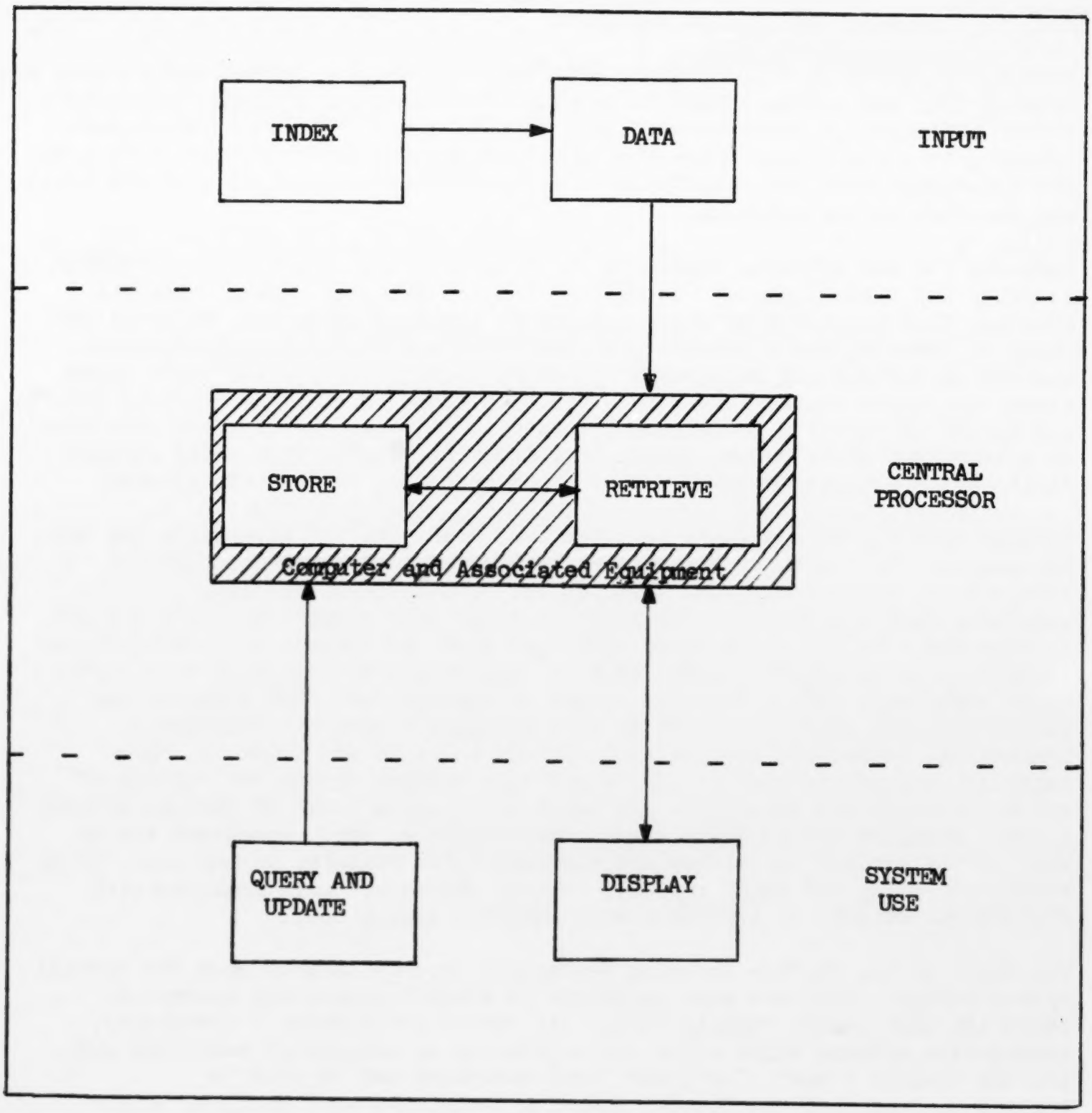

Figure 1. Conceptual Model of a Bibliographic Storage and Retrieval System

2.2 INDEXING AND SYSTEM INPUTS

Data to be stored in bibliographic retrieval systems are indexed before they are entered into the system. This is usually done manually, although some experimental automatic indexing schemes have been devised. In system development, indexing is a very significant step whose importance is frequently overlooked. All subsequent retrieval activities depend, to a large measure, upon the depth and accuracy of the indexing.

Indexing for the BCN will likely be based on the terms contained in the main headings and subheadings of the Medical Subject Headings (MESH) thesaurus. Although this thesaurus provides a standard indexing authority, accuracy and depth of indexing still require attention. In more and more applications, authors of serials and monographs are being required to provide index terms along with their manuscripts. This practice tends to minimize accuracy errors, but errors of omission reflecting the special and sometimes limited interests of authors may still occur. Also, some indexing schemes (not MESH) unnecessarily and sometimes severely restrict the number of index terms allowed.

Various indexing methods have been devised, each with its advantages and disadvantages. The indexing methods may be grouped into three categories: statistical, syntactical, and associative. Statistical techniques include weighting (and/or threshold settings) of terms with respect to their success in searches. Recent experiments have been directed towards using statistical techniques in an adaptive mode--that is, experience in searching with particular terms will add or subtract weight as appropriate. BCN planners may wish to include such techniques as more conclusive data are obtained. Syntactical techniques include such factors as links and roles or, more recently, content indicators, which serve to further refine the meaning of terms. Associative techniques are those which group terms of textual association. Examples are adjacent words, same sentence, etc. Sometimes two or more of the various techniques are combined. For example, in one case, in an experiment using the SMART system, content indicators were combined with statistical weights to provide a more powerful search tool.

The value of the various indexing techniques depends largely upon the overall system design. They are most important in a batch-processing operation, where the user cannot readily change his search formulation. Conversely, interactive systems which allow for multistage or sequential searching and provide Boolean ("and", "or", and "not) operators may be able to achieve much of what the special indexing techniques were designed to do. From a network viewpoint each indexing technique is likely to require its own thesaurus. While this can be handled insofar as hardware and software are concerned, it may create a problem for the user, since his search strategies and operational practices will depend partly upon the thesaurus. Thus, user at location A may find it difficult to interact with the index at location B, and vice versa. This is further discussed in Section 2.2 and 2.5.

One other input, the unit record, will generally have been specified in the system design. (Some systems allow the user to create a new unit record at will.) The unit record contains bibliographic data and the index terms, and may contain textual material. A major problem for the BCN concerning the unit record is the need for a set of minimum requirements for such records. A discussion of such requirements is contained in Section 3.2.1 on Selection Criteria.

2.3 SELECTED COMPUTER AND SOFTWARE CHARACTERISTICS

The analysis of selected computer and software characteristics is divided into three parts: File Characteristics, Operational Characteristics, and Performance Measures. By file characteristics we mean the structure and the relationship among several related aspects of files such as size, speed of search, updating, and accessing. By operational characteristics we refer to user-oriented capabilities such as Boolean operators and textual retrieval. (The dichotomy between file and operator characteristics is more a matter of convenience than necessity.) Performance measures include any techniques which serve as a basis for evaluating the degree and consistency with which a system does what it is supposed to do.

This analysis excludes such obviously important system features as the CPU, storage devices, I/O devices, core capacity, operating system, and the like. While these aspects are critical in understanding the operation of a system, they are generally reasonably well understood and require little explication. They are included in detailed comparisons of the systems in Section 3.2.3.

The factors contained in each of the three analyses which follow should be considered in terms of the several kinds of information usually retrieved from the unit record (such as title and author) and in terms of the differences that exist between on-line and batch operations. A structure for relating these many factors is presented in the matrix in Figure 2. The following discussions of these factors will be oriented to the on-line, interactive mode of operation; off-line, batch operations are discussed briefly in Section 2.4.2.

2.3.1 File Characteristics

The critical factors in this category include file size, speed of search, file structure, and the number and types of unit record elements by which the file can be accessed and updated. All of these are closely related.

In order to fully comprehend the critical factors associated with files, it is necessary to understand (1) the types of files found in a typical bibliographic data retrieval system; and (2) how documents are represented in them. These are detailed in the following discussion on file organization.

			OPERATIONAL MODE	
		UNIT RECORD CONTENTS / CRITICAL FACTORS	ON-LINE, INTERACTIVE a b c d	OFF-LINE, BATCH a b c d . . . .
BROAD CATEGORY	File Characteristics	A B C D E	(data)	(data)
	Operational Characteristics	A B C D	(data)	(data)

Figure 2. Matrix for Relating Critical Factors, Unit Record Contents, and Operational Mode

2.3.1.1 File Organization

Conceptually, bibliographic retrieval systems have three files:* the unit record file, the index file, and the postings file. The unit record provides a structured basis for storing those data which we consider important about each document. Ultimately, all bibliographic data must be retrieved from those records. The index file is a list of every distinct term on which we want to search and retrieve data. As a minimum this would contain the MESH main headings and subheadings and the names of all authors. It will undoubtedly contain much more and it may be divided into several index files. Each index term has a "pointer" associated with it. These pointers point to the postings file, which lists all documents that contain the particular index term. Processing a search request with these types of file structures is depicted in Figure 3 and explained below.

When the user enters a search statement, the computer searches the index file for a match. If a match is found, the pointer directs the program to the postings file, where all documents posted against the index term are listed. The program then goes to the addresses in the unit record file and retrieves either the unit records or the selected elements thereof, depending upon the design of the program and the commands issued to it for that particular case. Again, depending upon the computer program design, the system may cycle back from the postings file to the index file. This would be the case where Boolean operators are used to obtain the final postings list, i.e., the one that meets all Boolean requirements. The program may further process data obtained from the unit record file, for example, where certain qualifiers or logical operators are allowed.

In the illustration in Figure 3, assume a user wanted to search for documents by Aaron. This search request would result in three postings and retrieval of the unit records for documents 105, 382, and 916. If he had specified "Aaron and Dramamine," the system would have recycled to the index file after first finding the three postings for Aaron; in the second pass, only document 916 would have qualified (only the unit record at the intersection of Aaron and Dramamine in the postings file would have been retrieved). On the other hand, had he specified "Aaron or Dramamine," more than three unit records would probably have been found. In any case, the user might have qualified (modified) his request (for printing or output) by stating that he only wanted, say, title and date, if the system were so designed. Or he might have said that he only wanted those documents published since 1965; if document 916 has an earlier date, he would have found no postings for his query.

---

* Three files are the minimum required for an inverted or item-on-term file. This type of file structure is implicit for systems in the BCN. Although other file structures may be used, they are too limited in capability to be of interest to BCN planners.

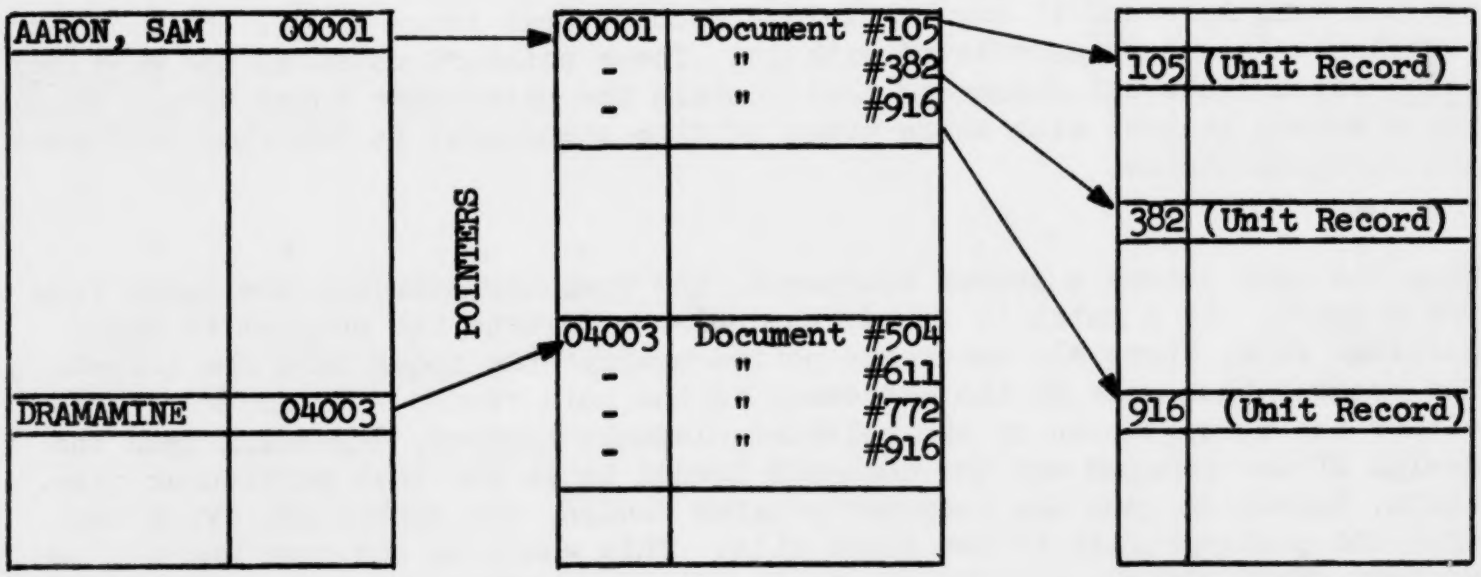

Figure 3. Relation of Files for Processing Search Statements

2.3.1.2 File Size

File size is fundamental in considering any storage and retrieval system. For one thing, it limits the amount of data that can be stored and the depth or richness to which the data can be indexed or otherwise tagged. Another point is that file size significantly affects costs in terms of the initial purchase price for the storage devices, their maintenance costs, and the costs of updating and purging the files. Obviously, file size can be increased by the simple expedient of adding more storage capacity, but this is limited by the number of storage units a computer can handle. Also, it usually means slower search speed and it may well affect other file characteristics. Indeed, speed of search and file size can, within limits, be traded against each other. With small data bases, speed of search is generally of no concern to the user, since computer speeds are so much greater than either the user or his I/O devices; with large data bases, however, search (access) time may become significant. In any event, the file size and speed of search are extremely critical--both in design and operations--to the cost of the computer system.

Size and speed are in some ways determined by file structure. For example, an inverted file, efficient and economical in terms of speed, imposes a burden in terms of size. File structure also dictates the way in which the file will be accessed, an important determinant of speed.

A review of quantitative data related to file characteristics for bibliographic storage and retrieval systems will facilitate understanding those characteristics as they are presented in the comparisons which follow. The size of the unit record file is directly proportionate to the number of entries and the characters per entry (which may be quite variable). As an example, 25,000 unit records (a modest collection) at 1,000 characters per entry (a plausible size) would require a storage capacity of 25×10^6 bytes. This approximates the capacity of one 2314 disk.

The size of the index file depends upon the number and length of index terms and the length of the pointers associated with each term. For programming reasons, irrelevant here, the entries in the index file are of fixed length. A typical entry for MESH main headings and subheadings would allow 50 characters for the term and the pointer. Thus 50 times the number of terms would indicate file size. For MESH this would be fairly constant--about 10,000 terms for any collection beyond a thousand or so documents. This is because the MESH vocabulary consists of about 10,000 terms, most of which would appear in the first few thousand documents indexed. Very few new index terms are likely to be encountered by adding a document after, say 2,000.

On the other hand, new author names will continuously appear as more documents are added, Similarly, if we were to include the title words in the index file, this would have an even greater effect on the file size. Hypothetical distribution curves indicating the number of items as a function of the number of documents for title words and author names are shown in Figure 4. Note that for the title words the most common words (a, and, the, for, by, on, in, etc.--about 1,000 in all) are suppressed. Since the title words are mostly MESH terms, that curve will reach an asymptote at about 10,000 terms, whereas the authors'-names portion of the file will continue to grow as new documents are added.

The postings file size is simply a function of the number of entries and their length. The length of entries is determined by the number of terms or items per document on which retrieval is allowed. For example if, in a given document, we permitted retrieval on 10 index terms, two authors, the language, and a document-type indicator, we would have 14 entries (10+2+1+1) in the file. Since 4 bytes will generally suffice for each entry, we require 56 bytes (14 x 4) for that document. A 100,000-document collection with 40 retrieval terms per document (a plausible number) would require 16×10^6 bytes of storage.

The combined size of all the files is a function of the number of documents, the number of index terms, and the ways in which searches may be accomplished. Since file size is generally fixed, it makes more sense to talk in terms of document capacity as a function of file size, search strategies, and index terms. This relationship is depicted for a hypothetical case in Table 1. The table is based on bibliographic citations that have been indexed with MESH terms. A large file size of 300×10^8 characters is assumed. The effects of six different search strategies and the index terms used are then shown.

Clear-text, full-text, or natural language searching is conceptually no different from what has already been described: it differs in degree rather than kind. In clear-text searching, each word or phrase to be potentially searched becomes an index term. The effect of doing this is illustrated in Table 1 with the addition of a 400-word abstract in which one can search English words (Search Strategy 6); the inherent problem is that it soon tends to overload the index file and/or core capacity. As long as the data base is small there is no problem; as the data base grows, the system has to search segments of the index, usually calling part of the file into core, searching it, holding those terms that match, searching the next segment, and so on. The effect is a significant increase in "swap time" or file search and access time. This raises serious doubts as to the cost-effectiveness of full-text searching with large data bases, especially since equally good or better retrieval may be obtained by using sophisticated techniques such as those discussed in Section 2.4.2.

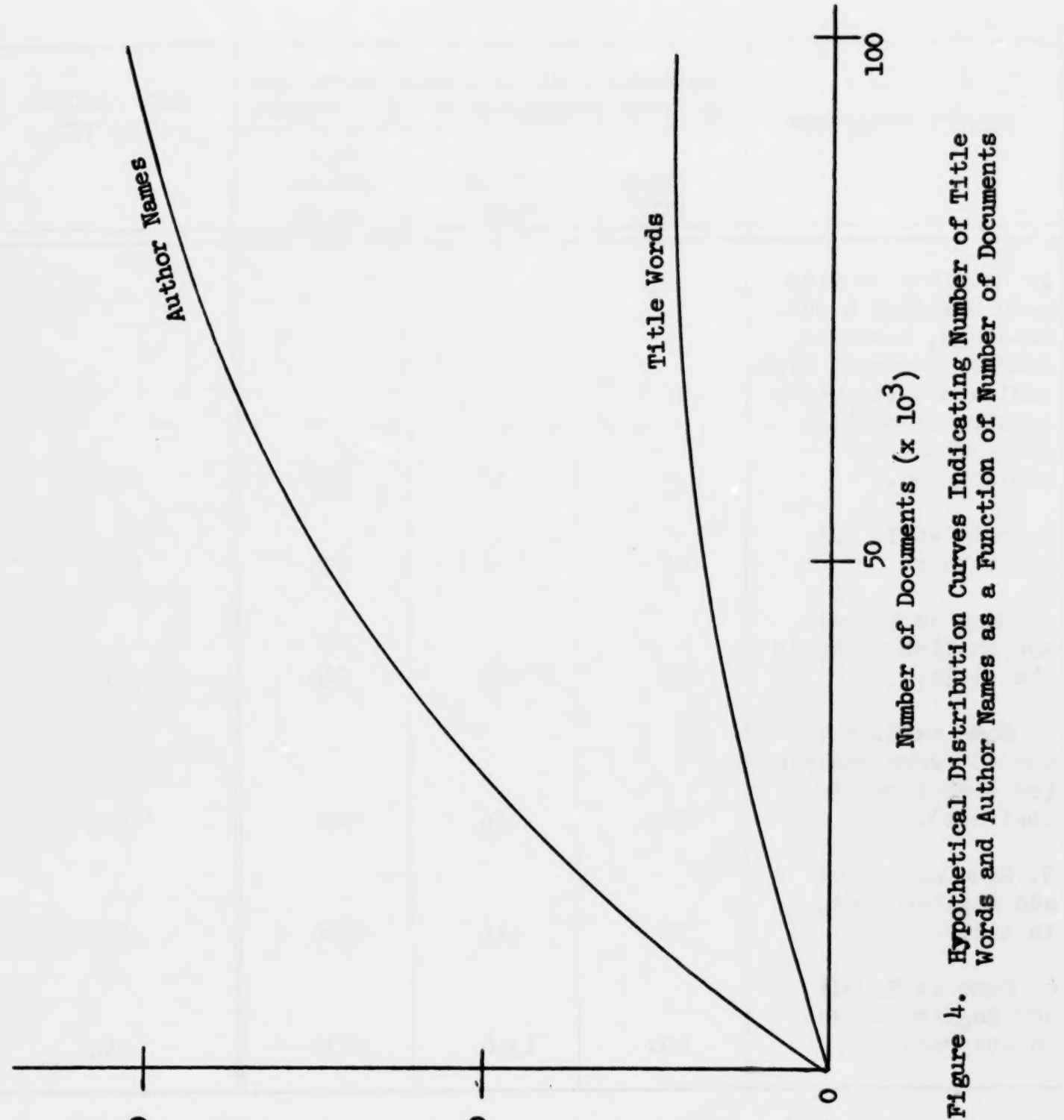

Figure 4. Hypothetical Distribution Curves Indicating Number of Title Words and Author Names as a Function of Number of Documents

Table 1. File Sizes and Unit Record Capacity for a Total File Capacity of 300×10^8 Bytes.

SEARCH STRATEGY*	MARGINAL COST OF ADDING DOCUMENTS IN BYTES (CHARACTERS) PER DOCUMENT			UNIT RECORDS CAPACITY
	INDEX FILE	POSTING FILE	UNIT RECORD FILE*	
1. Retrieve on MESH main headings & sub-headings, document source, document type indicator, language, country of publication, & date of publication.	0	52	630	444K
2. Same as 1. but add author.	25	56	630	423K
3. Same as 2. but add English words in the title.	75	116	630	365K
4. Same as 2. but add 400-word abstract (no search within abstract).	25	56	2630	110K
5. Same as 4. but add English words in title.	75	116	2630	106K
6. Same as 5. but add English words in abstract.	100	1316	2630	74K

*For strategies 1 through 3 inclusive the file contains the title, authors, MESH main headings and subheadings, document source, document type indicator, language, country of publication, and date of publication. For strategies 4 through 6 inclusive, the same elements plus a 400-word abstract are included.

On-line updating affects size in that it requires blank spaces in the several files for inserting new material. The amount of such space is arbitrary--a function of how much on-line updating is to be done. Generally, statistical decision rules are incorporated into a system design such that the size and quantity of updates per unit time are used to determine how many blank spaces, and of what size, to include. By careful analysis or by measuring the updates, their mean size and quantity can be determined. These data are used to refine the estimates of the update capacity required and optimize that capacity. Too many blank spaces are costly in terms of file size, whereas too few may preclude efficient operation or may require frequent recycling of the entire file. Note that structured files inhibit updating (the data must go to a special place), and random files facilitate updating (the data may go anywhere).

Internal computer-access time is very closely related to file size. It is determined by the time required to perform the several processes described in searching and going from file to file. Generally, the number of unit records will have little effect on access time; similarly, the number of index terms will have little effect. Each will contribute to a small fraction of a second per request regardless of file size. Conversely, the time required for the computer to get and process a postings record is very sensitive to the number of postings, the Boolean operators used, and the size of high-speed computer memory. For a simple search with less than 10,000 postings in the file, this time may well be considerably less than 1 second (the total of all parts being about 5 seconds* for a one-term retrieval). With 100,000 postings, however, this time becomes 120 seconds or 2 minutes. There is a general belief that user wait time in excess of 30 seconds is likely to be unacceptable.

One final but very significant point with respect to file size and speed of search is the size of core and its method of allocation in time-shared systems. As previously indicated, segmented searches, i.e., searches in which only part of the file is brought into core at one time, tend to be very slow. Thus the size of core relative to the file size is important. In time-shared systems this is further complicated by the fact that in some systems core is divided into segments assigned to each user. Such systems are core- or memory-shared rather than "time-shared," even though they are so called. In effect, this means that the user has a smaller core. The alternative mode is the one in which each user is swapped into and out of core periodically. In this scheme the user gets the entire core (except for a small portion required by the executive), but only for a fraction of the time. These systems are more literally time-shared. For files the size envisioned for the BCN, the second approach is considerably more effective.

*This assumes a time-shared computer system where the user gets about 1/20th of the time available and all files may be randomly accessed.

2.3.2 Operational Characteristics

The separation of operational from file characteristics is more apparent than real. This will become evident as the critical factors in this category are discussed. The operational characteristics we include in our discussion are retrieval using Boolean operators, the order dependence in retrieval, the textual retrieval capability, the on-line versus the off-line capabilities, and nonbibliographic retrieval.

Any sophisticated information storage and retrieval system will use Boolean operators. In particular, "and," "or," and "not" are required for the kind of search activities anticipated for the BCN. The system should also allow linking terms from different search statements or several request inputs. To facilitate users' operation, the system should be order independent, that is, there should not be any dependency upon a structured order for the input of commands, search statements, or other inputs. (Note that this is similar to the retrieval and output requirements in that they, too, are order independent.)

Textual retrieval was discussed briefly in conjunction with file characteristics. From an operational viewpoint, two aspects should be considered: search and retrieval (or input and output). The search aspect is concerned with the information stored in the index file. That is, in order to do textual searching each term on which one will search must be contained in the index file. Several problems become apparent. If the term is commonly used in many text entries (which is often the case), then the term must be linked in some manner with other term(s) so as to isolate it for a particular search. One technique for so doing is content dependency, where a generic term is linked with another term giving more precise meaning to the generic term. For example, the term "chemotherapy" might be so frequent in a medical data base that it would not be meaningful unless linked somehow with another term such as a disease name. Thus chemotherapy for leukemia, for syphillis, or for schizophrenia are very meaningful and serve to limit an otherwise general term. Several other schemes exist for linking terms. All require that the files be structured in certain ways. Full-text searching does not always result in more effective retrieval, but it always adds to the cost. Furthermore, the use of Boolean operators and linking may ultimately be considered the operational equivalent of full-text retrieval.

Textual retrieval or output depends upon the content of the unit record file. Obviously, no text retrieval can be performed except on text that exists in text form in the unit record or an additional file.* File size will tend to limit the amount of text in the unit records. As indicated in Table 1, extensive text storage soon becomes very expensive. If extensive text is required, the user will normally want to go to the original document; rarely would the user tolerate on-line printing of extensive textual messages. But an abstract limited to

*The text may be stored elsewhere with its address or location stored in the unit record.

a few hundred words might sometimes be desired (even here there is a noticeable effect on file size--see Table 1 and accompanying discussion) and the ability to print such material, on-line or off-line, seems desirable for the BCN. Requirements for textual search or retrieval are further considered in the discussion about textual search and retrieval in Section 3.2.1, Selection Criteria.

Batch processing, off-line systems have been designed and used for bibliographic retrieval and should therefore be mentioned. Batch operations may provide an important adjunct to on-line operations. This is especially true where, as just described, extensive text retrieval is desired, or where extensive listings such as special bibliographies are desired; some batch systems (e.g., the MARC IV) have provided retrospective search as well. The functions provided by batch systems are generally also available with interactive on-line systems. That is, interactive systems typically have an off-line capability that allows the user to specify those outputs he wishes produced off-line. Furthermore, the user can sample the output and determine, a priori, that the format and content are as desired.

2.3.3 Performance Measures

With the development of storage and retrieval systems has come the development of performance measures for such systems. Attempts have been made to establish a single figure of merit by which to evaluate bibliographic systems. As yet, no standard has been generally accepted. One well-documented measure is the precision/recall (P/R) ratio. Precision measures the relevance of the items that are retrieved, whereas recall measures the number of items of the total relevant set retrieved. Precision/recall data have been reported for the NLM, and intensive studies of such data are available for a limited number of systems. However, no such data are available for most of the systems compared in this study.

One major drawback with P/R is that it fails to isolate those elements of the system about which we may want performance information. In fact, P/R is almost exclusively a function of the original indexing and user search strategy and sophistication. It therefore tells us almost nothing about the effects of file organization, on-line versus off-line retrieval, or other aspects of the system.

Another potential performance measure is a relevance measure which evaluates the degree of relevance of a retrieved document. Relevance has been used primarily as a feedback mechanism rather than a performance measure.

An area of performance which has not as yet been systematically measured in any thorough way is a comparison of on-line versus off-line retrieval with particular emphasis on the effect of the interaction between the human operator and the system. Specifically, on-line interactive users who receive immediate feedback from the system are likely to receive many cues which will lead them in directions not originally anticipated; this results in an adaptive or heuristic quality that further enhances searching. Users adjust their search strategy in terms of immediate relevance judgments. This serendipitous effect of interactive systems has not been systematically measured and yet may constitute a very important element of system performance.

2.4 USER CONSIDERATIONS

The third major functional area of the system model presented in Figure 1 is that of system use. The design and capabilities of the system are intended to satisfy user requirements. Equally important, user knowledge, attitude, and behavior will determine the performance and utility of the system. The user is constrained by the I/O device or devices available, as well as by the many system and operating characteristics already described. He is further constrained by the user language, although this constraint should be minimal in any well-designed system; that is, the language should be sufficiently close to natural language and the formats sufficiently simple so that minimal learning is required for system use. In this connection the system should be designed such that user errors (either typing or input errors) will not degrade system performance, interfere with other users, or otherwise have detrimental effects upon the programs or equipment. Also, the system should be designed to ensure that the user will be able to recognize and easily correct any errors.

The search strategies available to the user will, to a large degree, determine how effectively he uses the system. The richer the set of such strategies, the more power the system will have--but the larger the burden in file size, as previously discussed.

The system should be easy to use. That is, it should not require any extensive training or special skills. For example, systems using remote typewriters or teletypewriters typically require only a few characters be typed by the user. This, combined with the error detection and correction capability and a simple input language, creates an easy-to-use system. However, the novice will require some training. This training will be greatly enhanced if there are built-in tutorial messages and instructions to the user. By providing user option in selecting cascaded levels of instructions, cues, or explanations, the system is readily adaptable to the user's capability. Thus an experienced user is presented only minimal cues, whereas the novice is presented detailed instructions and may request further explanation.

Another important user consideration is response time. Users who previously had to wait overnight or even days to get information in response to their requests become very impatient when forced to wait more than 30 seconds at a terminal. Different systems operate differently in this respect: some (dedicated) guarantee fast feedback as long as the system is operational; others (time-shared) have delays in response which may in part be a function of the number of users and/or the size of their programs. While not an essential requirement in the design of the BCN, response time should be considered and, where practicable, limited to 30 seconds.

Finally, a number of specific capabilities that will enhance the ease of use and increase user acceptability (and consequently system utilization) follow. The user should be able to:

1. Enter any command at any time, i.e., there should not be any ordering or sequential requirements on the inputs.

2. Select which records to print and specify which unit record elements to print (or delete) in a retrieval.

3. Qualify retrievals on the basis of factors he considers significant, e.g., publication date, publication language.

4. Print information on line (i.e., retrieve hard copy) or, at his discretion, print information off line--especially lengthy reports such as demand bibliographies.

5. Have unrestricted use of Boolean operators and be allowed many index terms--say 20 or more--per search statement.

6. Link search statements and/or commands with little if any restriction as to their length; and truncate search statements, as desired.

7. Receive unambiguous cues as to when to input requests or commands.

8. Nest his inputs or search statements to at least five levels.

9. Input any data in any form recognized by the program (e.g., symbolic or coded, short or mnemonic, or full name) and mix such forms at will.

10. Obtain information about adjacent or similar words in the thesaurus.

11. Formulate search statements on any element as it may be contained in the files without having to label that element. If the element may be more than one type or have multiple meanings (for example, the term Taylor might be an author's name or a publisher), then the program should indicate that fact, indicate the choices, and ask the user to specify which was intended.

12. Be protected against unintentional destruction of the files, other users' programs, or the system programs are concerned. For example, if the user enters any command which has a significant effect upon his status (e.g., disconnects his terminal) or his files (e.g., deletes an item), the program should double check, that is, require positive control that this is indeed what the user intended, before allowing the user to execute the command. Furthermore, only authorized users should have access to the files or system programs.

Other features, such as the ability to obtain thesaurus segments on-line and to search on terms subordinate to any index term for hierarchically organized data bases, provide extremely powerful user capabilities.

2.5 NETWORK FACTORS

While network planning is beyond the scope of this report, a brief discussion of network factors as they relate to the analysis of retrieval systems seems warranted. We take as our basis for discussion the concepts presented in NLM's initial Technical Development Plan for the Biomedical Communications Network. That plan identified several BCN components including a library component and a data processing and data transmission facilities support component. Our concern is with certain applications and operational aspects of these components, rather than with their technical aspects. The NLM plan indicated great insight into the technical problems such as equipment and data rate compatibility. These problems are not discussed here.

The initial BCN plan calls for a multi-level network with four levels or layers. The Center for Biomedical Communications (CBC) is the central point and many hundreds of local libraries are the points of contact for the majority of users. Intermediate levels will serve as the interface between the local libraries and the CBC. Intricate communications links will exist vertically and horizontally throughout the network. Implicit in this design is an operational protocol whose essential steps are:

1. A user makes a request at a local library for information not contained there.

2. The request is entered into the network by someone (perhaps the user, more likely a librarian or information specialist) by means of a remote terminal.

3. The remote terminal entry causes computer files at some other location to be interrogated. This may be totally on line,* off line, or mixed. The files may or may not be "in residence" on some computer storage device.

4. If the transaction is on line, the information requested will be printed at the remote terminal and will be immediately available for use. If not, the response will be sent to the user later, perhaps by mail.

Presumably, a significant portion of the transactions will be on-line. The remainder of this section is concerned with problems relating such transactions and with several aspects of bibliographic retrieval.

---

*Although the plan does not allow for remote terminal users at local terminals to link directly to a computer file, interface devices and/or procedures could provide essentially on-line operations.

In terms of the three major components into which retrieval systems were earlier divided (inputs, central processor, and system use), the one that received the least attention in the BCN plan, insofar as networking problems are concerned, was system use. Some of the problems concerned with system use stem from the fact that the user will not be able to communicate directly with the computer file. Rather, he will work through an intermediary. The likely result is suboptimization of retrieval due to the lack of immediate feedback and relevance information. Also the user will not be able to exploit the search strategies to their fullest.

With respect to inputs, few problems are anticipated. One area of caution is required if files are created by the secondary centers. Vigorous control will have to be exercised so that indexing is consistent among the several centers. Failure to do so is likely to result in different file characteristics and thereby degrade network performance.

With respect to the computers and their associated software, any potential problems of compatibility are mitigated by the fact that various interfacing, conversion, and transformation devices and techniques may be used to compensate for incompatibilities. The problem becomes one of minimizing the integration requirements without compromising the unique capabilities desired by each of the subordinate centers. The solutions must ultimately be in terms of costs, benefits, and efficiencies. In this respect there are potential trade-offs between degree of integration and degree conformity to a set of standards.

Finally, there are problems related to incorporating elements not specifically designed for the BCN into the network that are discussed in Section 4.3.

3. SYSTEM COMPARISONS

Recent years have witnessed a proliferation of information storage and retrieval systems. A few have been specifically designed for bibliographic data retrieval. Most, however, are general-purpose data management or storage and retrieval systems which may be used, although less efficiently, for bibliographic search and retrieval. There are hundreds of such general-purpose systems and several more emerge each month.

Classifying storage and retrieval systems into general-purpose and bibliographic systems is convenient for comparing the systems. Although prime interest centers on bibliographic systems, general-purpose systems cannot be totally dismissed, since they are more readily available and often have performed, or have the potential for performing, bibliographic search. In such cases special programs are written for the general-purpose system. These programs usually call upon certain functions of the general-purpose system--such as file manipulation--to perform some of the retrieval tasks. Indexes and other files are often structured in the format of the general-purpose system.

Since the BCN might make use of one or more general-purpose systems as an adjunct to the network, it is necessary to consider them. However, the comparisons are considerably less detailed than those performed for bibliographic systems. Since a major purpose is to ascertain implications for BCN design from existing systems, reviewing the features of a set of representative or typical systems will accomplish the objective. Accordingly, of about 150 systems initially reviewed, 26 general-purpose and 11 bibliographic systems were selected for comparison. The basis for selection was twofold: systems were selected in terms of criteria established for each of two categories, and only those systems for which data were readily available from open sources were included. Additionally, the author arbitrarily included or excluded some systems, depending upon their perceived relevance for BCN planning. For example, systems in which bibliographic retrieval is a central or important function were included; others were eliminated because they were too specialized, too old, or lacked readily available information.*

*Several foreign systems that seem to parallel developments in the United States are not included in this report. The reader who desires more information should see the listings in Current Research and Development in Scientific Documentation, No. 15, National Science Foundation, NSF-69-8.

The general-purpose systems selected for comparison are listed and compared in Section 3.1. Section 3.2 contains the selection criteria for choosing the bibliographic systems to be compared as well as the detailed comparisons of those systems. A summary of the comparisons is contained in Section 3.3.

3.1 GENERAL-PURPOSE DATA MANAGEMENT SYSTEMS

The 26 general-purpose systems listed in this section vary in size, capability, specificity or generality, applications, and technical features. All are on-line systems. (Thus, such well known systems as MARK IV and MANAGE were excluded.) All may be operated remotely with at least one of the more common remote terminals such as a teletypewriter or a 2741 typewriter. Finally, all are capable of bibliographic search and retrieval--usually with some modification--although few have been used for that purpose.

Library technical processing programs are sometimes available in conjunction with a data management system. These programs perform such functions as acquisitions, cataloging, serials control, and circulation control. An example is SDC's ALPS (Automated Library Processing Services). These functions may be of interest in BCN planning in that they provide services ancillary to information storage and retrieval. However, such features were beyond the scope of this report and are not considered in the comparisons.

One system not included in the comparisons is that associated with MIT's Project MAC. The system, CTSS, and it's successor MULTICS, is used by MIT as an on-campus (and limited public) system. The system itself is of little interest but Project MAC gave rise to important concepts in the field of general-purpose data management systems. Those concepts have been reflected throughout this report. Project MAC, in association with the American Institute of Physics, is responsible for developing the Technical Information Project (TIP). Project INTREX, also associated with MIT and Project MAC, is an important experiment in the area of information handling. Both TIP and INTREX are discussed in Section 3.2.

The comparisons in this section are limited to general statements of the systems' applications and their major technical features. This was felt to be adequate for determining their implications for the BCN. The several general-purpose systems are compared in Table 2. All of the systems listed are operational.

(page 24 blank)

Several comments will aid in interpreting the data contained in Table 2. The dynamic nature of this field--systems are being created, modified, and abandoned at a fairly high rate--makes much of the information perishable. Therefore, the data should be interpreted as representing the current situation, but are likely to change rapidly. Furthermore, a number of the systems exist in two or more configurations with different capacities; the data shown are for the typical case. The CPU listed in the table is the smallest of the series on which the system is operational; the systems generally can be run on any larger computer of the same series. The number of simultaneous on-line users is often misleading. Published data are usually for the number of ports or channels available, and the system could theoretically accommodate that number of users. In practice, however, some of the larger numbers are not feasible, since various system capacities are soon exceeded. The user/core allocation refers to the amount of core allocated when the user's program is run. "All" means all available core, not all core, since the executive or other programs may permanently reside in core. The amount used by the resident programs varies but generally would not run above 20% of core capacity. The mass storage capacity is the on-line capacity and is the upper limit that the user could theoretically have available. In reality the capacity must be divided among all users and one user would rarely have it all available to him except in a dedicated mode. The term "virtually no limit" indicates that no restrictions are placed on the size of the user file. However, the user file could not exceed the mass storage capacity and it would generally be considerably less than that capacity. Most systems will operate with a wide variety of remote terminals. The ones listed are those commonly used.

Table 2. List and Brief Description of Selected General-Purpose Systems

SHORT NAME	NAME	BRIEF DESCRIPTION AND APPLICATIONS E--Educational PD--Program Debugging F--Fiscal S--Scheduling I--Inventory SE--Scientific/ Engineering PA--Personnel/ Administrative T--Text Edit	CPU	NO. OF SIMUL-TA-NEOUS ON-LINE USERS	C S C C
ADAM	Advanced Data Management System	General-purpose file management system. Provides multi-level, tree structured hierarchy. Produced primarily as a laboratory design tool to model information processing systems.	IBM 7030	7	6 V (
---	Allen-Babcock T.S. System	Commercially available TS system with conversational language (RUSH--a variant of P/L 1) with applications packages for all eight areas.	360/50	120	
---	BB&N Information Storage & Retrieval System	A time-shared data management system for storing and retrieving data files including medical records and document catalog information.	PDP-7/8 PDP-10	64	2 wo (re (
---	CEIR T.S. System	Commercially available TS system using ALGOL and BASIC as the conversational language. Applications above, except PA, I and limited in F.	GE 235 & GE 425	40	

*The term all indicates that the user has access to the total core except that portion reserved for the executive or other resident programs.

Systems

ABBREVIATIONS

Y--Yes
N--No
P--Planned
VNL--Virtually No Limit
TS--Time-Shared
A--Available

TECHNICAL FEATURES							DEVELOPED BY
CPU	NO. OF SIMUL-TA-NEOUS ON-LINE USERS	CORE STORE CAPA-CITY	USER/ CORE ALLO-CATION*	MASS STORE CAPA-CITY (ON-LINE)	USER FILE CAPA-CITY	COMMON-LY USED REMOTE TERMI-NALS	
IBM 7030	7	65K words (64 bit word)	all	$4x10^6$ words	VNL	DDI 2741 TTY	MITRE Corp.
360/50	120	460K Bytes	32 K fore-ground 250K back-ground	600×10^6 Bytes	40×10^6 Bytes (link-able)	2741 TTY DATEL DURA 1041	Allen Babcock Computing Co.
PDP-7/8 PDP-10	64	24K 64bit words (8 & 10 respect) (36 bit words)	all	50×10^6 char	32K per file, no limit on number of bits	TTY CALCOMP plotter	Bolt, Beranak & Newman
GE 235 & GE 425	40	16K & 32K	all	16×10^6 & 32×10^6 char.	127K char.	TTY Friden 7100	CEIR

xcept that

Table 2. List and Brief Description of Selected General-Purpose Systems (cont.)

SHORT NAME	NAME	BRIEF DESCRIPTION AND APPLICATIONS E--Educational PD--Program Debugging F--Fiscal S--Scheduling I--Inventory SE--Scientific/ Engineering PA--Personnel/ Administrative T--Text Edit	CPU	NO. OF SIMULTANEOUS ON-LINE USERS
---	Computer Network Corp. T.S. System	Commercially available TS system. Applications include F, I, PD, SE, T. Conversational languages are FORTRAN IV, BASIC, ALGOL, COBOL.	B 5500	40
---	Computer Sciences Corp. System	Commercially available TS system. All applications except F, I, and limited in administrative. Conversational language is a variant of BASIC.	UNIVAC 1108	50
---	Computer Sharing T.S. System	Commercially available general-purpose TS system. Specific applications unknown. Conversational languages are QED, BASIC, CAL, DDT, FORTRAN IV.	SDC 940	40
---	Computer Time Sharing T.S. System	Commercially available TS systems limited PD and SE. Uses BASIC, EDIT, FORTRAN II, & DEBUG as conversational languages.	CDC 3300	100
---	COM-SHARE	Commercially available TS system, with all applications listed. Conversational languages are BASIC and FORTRAN IV.	SDS 940	44
---	Control Data Corp. T.S. System	Commercially available TS systems. Applications are accounting, personnel, SE, and PD. Conversational languages and FORTRAN IV AND BASIC.	CDC 3300 6600	256
---	Data Network Corp. T.S. System	Commercially available TS system. All applications except E, I, and personnel. Uses FORTRAN II & IV, QED, BASIC, and CAL for conversation.	GE 420 (360/67 being installed)	40

ation
-4421

stems (cont.)

TECHNICAL FEATURES							DEVELOPED BY
CPU	NO. OF SIMUL-TA-NEOUS ON-LINE USERS	CORE STORE CAPA-CITY	USER/ CORE ALLO-CATION	MASS STORE CAPA-CITY (ON-LINE)	USER FILE CAPA-CITY	COMMON-LY USED REMOTE TERMI-NALS	
B 5500	40	32K words	all	10×10^6 char.	VNL	TTY CCI Friden	Computer Network Corp.
UNI-VAC 1108	50	131K words	all	157×10^6 char.	VNL	2741 TTY U-D CT 2000 U-11D4	Computer Sciences Corp.
SDC 940	40	32K words	No more than 3 files open at one time	67×10^6 char.	VNL	TTY Incre-mental plotter	Computer Sharing Inc.
CDC 3300	100	65K words	all	133×10^6 char. (&4×10^6 drum)	VNL	TTY	Computer Time Sharing Inc.
SDS 940	44	64K words	16K (24 bit) words	70×10^6 (mixed storage media)	VNL	TTY Friden 7100	Comshare Inc.
CDC 3300 6600	256	131K words	10K words	64K 6 bit words	no data files	TTY	Control Data Corp.
GE 420 (360/67 being instal-led)	40	64K words	all	67×10^6 char.	57K char. for 40 file	TTY Cal Comp Plotter Friden 7100 CCD CRT	Data Network Corp.

Table 2. List and Brief Description of Selected General-Purpose Systems (cont.)

SHORT NAME	NAME	BRIEF DESCRIPTION AND APPLICATIONS E--Educational PD--Program Debugging F--Fiscal S--Scheduling I--Inventory SE--Scientific/ Engineering PA--Personnel/ Administrative T--Text Edit	CPU	NO. SIM TA NEO ON-L USE
---	Direct Access Computer Corp. T.S. System	Commercially available TS system. All applications. Conversational languages are BASIC, FORTRAN IV, COBOL, ALGOL.	B 5500	48
---	Eli Time Sharing T.S. System	Commercially available TS system. All Applications except educational and administrative. Conversational languages are BASIC, ALGOL, COBOL, FORTRAN IV.	B5550	48
GIS	Generalized Information System	General data management system which can interrogate and maintain arbitrary user files in response to unstructured and unanticipated requests. Files transferable with others created under OS 360.	360/40	--
MARK II	General Electric MARK II System	Commercially available TS system. All applications. Conversational languages are BASIC, ALGOL, FORTRAN.	GE 635	200
---	IBM CALL/360	Commercially available TS system. Applications include SE, PD, E., plus simulation. Uses BASIC as conversational language. FORTRAN and PL/1 available.	360/50	not ava abl
---	ITT Data Services T.S. System	Commercially available TS system. Applications include all except E and limited in F. Conversational languages are FORTRAN, COBOL, BAL, and BASIC.	360/65	200

TECHNICAL FEATURES						DEVELOPED BY
NO. OF SIMUL-TA-NEOUS ON-LINE USERS	CORE STORE CAPA-CITY	USER/ CORE ALLO-CATION	MASS STORE CAPA-CITY (ON-LINE)	USER FILE CAPA-CITY	COMMON-LY USED REMOTE TERMI-NALS	
48	32K 48 bit words	all	96×10^6 char. (10 discs)	VNL	TTY	Direct Access Computer Corp.
48	32K (48 bit words)	all	150×10^6 char.	VNL	TTY	Eli Time Sharing Inc.
--	512K bytes	all	81×10^6 char. (min.)	VNL	TTY 2741	IBM
200	96K words (36 bit)	all	100×10^6 char.	VNL	TTY F - 7100 CAL COMP Plotter	General Electric Co.
not avail-able	512K bytes	All	232×10^6 bytes	VNL	2741 TTY	Service Bureau Corp.
200	512K bytes	61K bytes	205×10^6 bytes (&7.25×10^6 slow)	VNL	2741 1050 TTY F - 7100	ITT Data Services Inc.

Table 2. List and Brief Description of Selected General-Purpose Systems (cont.)

SHORT NAME	NAME	BRIEF DESCRIPTION AND APPLICATIONS E--Educational PD--Program Debugging F--Fiscal S--Scheduling I--Inventory SE--Scientific/ Engineering PA--Personnel/ Administrative T--Text Edit	CPU	NO. C SIMUL TA- NEOUS ON-LIN USERS
---	Keydata & Adams Associates T.S. System	Commercially available TS system. Applications include F, I, PD, and S. Conversational language is KOP-3.	UNI-VAC 491	200
---	McDonnell Automation T.S. System	Commercially available TS system. All applications. Uses BASIC and FORTRAN for conversational language.	GE 420	30
---	On-line T.S. System	Commercially available TS system. Applications except PA and limited in F. Conversational languages include BASIC, FORTRAN IV, MACRO, LISP, AID.	PDP 10/8	64
---	Philco-Ford Corp. T.S. System	Commercially available TS system. All applications. Conversational languages are BASIC, FORTRAN, ALGOL.	GE235 P-F 102 (2 ea)	192
---	Rapidata T.S. System	Commercially available TS system. All applications except PA. Conversational languages are BASIC and FORTRAN IV.	GE 435	44
---	Remote Computing Corp. T.S. System	Commercially available TS system. All applications present or planned. Conversational languages are COBOL, BASIC, ALGOL, FORTRAN.	B 5500	48
TS/DMS	Time Shared Data Management System	Commercially available TS system. Consists of 4 major elements: a TS executive, a general purpose data management system, special purpose data management systems, and a programmer's package. Performs all applications. User oriented conversational language.	360/67	40

Corporation
TM-4421

stems (cont.)

TECHNICAL FEATURES							DEVELOPED BY
CPU	NO. OF SIMUL-TA-NEOUS ON-LINE USERS	CORE STORE CAPA-CITY	USER/ CORE ALLO-CATION	MASS STORE CAPA-CITY (ON-LINE)	USER FILE CAPA-CITY	COMMON-LY USED REMOTE TERMI-NALS	
UNI-VAC 491	200	32K words	All	131 x 10^6 char.	13 x 10^6 words	TTY F - 7100	Keydata & Adams Assoc.
GE 420	30	32K words	18K words	23.5 x 10^6 char.	130K char. (8bits)	TTY F - 7100	McDonnel Auto-mation
PDP /8	64	64K (36 bit) words	All	52x10^6 char.	2 discs 2.6x10^6 for swap	TTYDatel 2741 F - 7100	On-Line Systems Inc.
GE235 P-F 102 (2 ea)	192	65K words	All	216x10^6 ch. + 28 x 10^6ch. (slow)	VNL	TTY	Philco-Ford Corp.
GE 435	44	64K words	All	64 x 10^6 char.	VNL	TTY F - 7100 2741 Cal Comp	Rapidata Inc.
B 5500	48	33K words (48bits)	All	128x10^6 char.	VNL	TTY	Remote Computing Inc.
360/67	40	512K bytes	All	432x10^6 char.	VNL	TTY 2741	System Develop-ment Corp.

Table 2. List and Brief Description of Selected General-Purpose Systems (cont.)

SHORT NAME	NAME	BRIEF DESCRIPTION AND APPLICATIONS E--Educational PD--Program Debugging F--Fiscal S--Scheduling I--Inventory SE--Scientific/ Engineering PA--Personnel/ Administrative T--Text Edit	CPU	NO. OF SIMULTANEOUS ON-LINE USERS
---	Tymeshare System	Commercially available TS system. All applications. Conversational languages are SUPER BASIC, FORTRAN IV, EDITOR, DDT, CAL.	SDS940	42
---	U.S. Time Sharing System	Commercially available TS system. All applications. Conversational language is PLUS 1.	360/65	60

e Systems (cont.)

	TECHNICAL FEATURES							DEVELOPED BY
	CPU	NO. OF SIMUL-TA-NEOUS ON-LINE USERS	CORE STORE CAPA-CITY	USER/ CORE ALLO-CATION	MASS STORE CAPA-CITY (ON-LINE)	USER FILE CAPA-CITY	COMMON-LY USED REMOTE TERMI-NALS	
	SDS940	42	192K bytes	All	65×10^6 char. ($+6 \times 10^6$ char. slow)	48×10^6 char.	2741 TTY CalComp Plotter CCI-33 K - CRT	Tymeshare Inc.
	360/65	60	512K bytes	All	223×10^6 bytes	VNL	2741 TTY 1050 2260 2780	U.S. Time-Sharing Inc.

3.2 BIBLIOGRAPHIC DATA RETRIEVAL SYSTEMS

Relatively few systems have been designed specifically as bibliographic data retrieval systems; hopefully, all of the significant ones (in terms of BCN planning) are included in this section. Some relatively small or relatively unknown systems are not included, since data were not readily available. Others were excluded because of their limited and specialized application or the fact that they were transitional and no longer in operation. Auerbach's IM-1, a general-purpose system, is included as an example of a general-purpose system that has been used for bibliographic retrieval. Although similar to ORBIT II, ELHILL was included as a separate system for comparison since it was specifically designed for NLM use in BCN development.

Two projects not included in the detailed comparisons, since they are research projects rather than systems, are TIP and INTREX. TIP (Technical Information Project), developed by MIT and the American Institute of Physics, is a project concerned with storing and retrieving physics data. It is used via the Project MAC time-sharing system. Although a significant project in its area of application, TIP has little relevance for BCN planning. INTREX (Information Transfer Experiment), an extremely ambitious undertaking, is concerned with basic principles of information exchange. For example, a computer-based augmented catalog system is now being developed, and the work on integrating the TEXT ACCESS program with the INTREX display console was completed in 1969. This latter provides users access to the full text of documents in the INTREX data base which may then be photographed and stored on microfiche. Hard copy retrieval is then possible. The size, complexity, and importance of INTREX suggest that NLM continue to monitor its progress. The techniques developed will have impact on any retrieval systems or networks developed in 1970's.

The criteria for including systems for detailed comparisons are described in Section 3.2.1. The system features that are compared are described in Sections 3.2.2 and 3.2.3. The comparisons are sometimes limited to noting the presence or absence of features, and do not necessarily include details as to how those features are implemented. This factor needs to be considered in interpreting the comparisons.

3.2.1 Selection Criteria

The five criteria that were used to select bibliographic retrieval systems for detailed comparison constitute a set of basic requirements for BCN planning. These requirements include the conceptual and operational features that any system must contain in order to be part of or effectively contribute to the design of the BCN. Thus, such basic requirements as "the system must provide for on-line interactive remote terminals" or "the system must be compatible with the MEDLARS MESH" are included, although the type of terminals and the question of whether the system has ever been used for MESH are of no concern here. Some systems were included despite certain shortcomings with respect to the criteria. These systems, however, potentially fulfilled the criteria and doing this is planned. Nevertheless, the criteria are the minimal acceptable set of requirements for BCN planning.

1. The system must provide for on-line interactive remote terminals.

Implicit in a network design is the ability of network users to input queries and rapidly retrieve information based on those queries. To provide fast response time--particularly at great distance--requires some kind of interactive remote terminal. To satisfy this requirement, any of the commonly used terminals such as typewriters, teletypewriters, or CRT displays might be suitable. There is, however, a clear preference for the ability to have hard copy of data retrieved, at least some of the time. No conclusive data were found concerning this point. While some users argue that video displays allow quicker and perhaps more extensive searching (especially when one is not sure what he is searching for), almost all users argue that the ability to produce some hard copy at the terminal is mandatory. This is further suggested by the fact that the terminal user is often not the same as the ultimate user. Therefore the terminal user often needs the copy to present to the ultimate user. This would suggest typewriter or teletypewriter, although other combinations such as a CRT display* and teletypewriter are acceptable and must be considered from a cost-benefit perspective.

2. The system must be able to store, search, and retrieve bibliographic data.

This reflects one of the major purposes for which the BCN is being developed. A standard is yet to be established for the bibliographic unit record that must be stored. However, as a minimum, the record should contain author(s), title, publication date, serial reference (as appropriate), and index terms--the MESH main headings and subheadings. Ideally, the unit record should also contain reference and/or acquisition numbers and information concerning the language of publication. Other data that are useful are the country of publication and the document source.

3. The system should be user oriented.

Several important features are implied here, although only the first three are mandatory criteria for including a system in this section. First, the queries should be easy to formulate in natural or near-natural language. Second, the responses should be readily understood and also in natural language (or an easily understandable shorthand). Third, the system should tolerate errors as well as inform the user as to the nature of his errors.

*In this connection we should distinguish between systems which display video or graphic data versus those that use CRT's as an alphanumeric display device. The former require wideband video communications channels and, because of this communications requirement, are extremely expensive to operate from remote terminals. Their need in the BCN has not been demonstrated. CRT's which display alphanumerics, such as those used in the CCI unit, may be used with the same telephone and teletype lines that are used for the other remote terminals.

It should also provide him with a basis for taking corrective action. Fourth, the system should be sufficiently simple to learn and operate (or contain sufficient tutorial information) so that a novice can use it successfully (although not necessarily with maximum efficiency) with not more than one hour of instruction. Fifth, the system should allow the user to select minimal cues or detailed instructions, depending upon his level of familiarity. Other desirable niceties are feedback concerning the relevance of the data retrieved and feedback concerning alternate actions that the user might take at any given point in operating this system.

4. The system must be compatible with a 360/67 or equivalent computer.

This requirement is based on the fact that the IBM 360 series is most frequently found among the systems that are likely candidates for networking and because NLM presently uses a 360/50. The 360/67 was selected because of the large core capacity that can be associated with it and of users that may time-share the system. This does not preclude the use of other than IBM computers, but suggests that if any are used, they and their interfacing devices be compatible with the 360/67.*

5. Files should be organized on a random-access basis and the relationship of the file size to the fast core size should be such that it is not necessary to perform sequential searches on segments of the file.

Time penalties and poor cost-effective use of the system would result were this requirement not met. Some of the systems included had random-access capability but required limited sequential switching, since the file size exceeded the storage device capacity.

3.2.2 Overview

Table 3 contains an overview of the systems selected for detailed comparisons. The contents of the table are:

Developer. Specifies the organization that developed and/or currently operates the system.

Brief Description and Other Features. (Self-explanatory.)

Bibliographic Applications. Indicates if used for medical or other special bibliographic retrieval purposes.

Availability. Indicates whether the system is publicly available or limited to some special group.

*In reality, costs rather than technical factors ultimately dictate compatibility. With appropriate interfacing devices any hardware devices may be made compatible with any other. The question becomes, does it pay?

(page 39 blank)

With respect to Table 3 (and Tables 4 through 7), several systems have had their names changed, have been incorporated into larger systems, or exist in two or more versions. Where multiple versions exist, they were considered equivalent (i.e., treated as one) despite their differences.* When, for whatever reason, more than one name occurred, an arbitrary choice, usually reflecting the more recent, operational, or more advanced version, was used in the tables. The "equivalent" names are:

Name Used in Tables	Other Names by Which Known or Included Systems
RECON	DIALOG
BALLOTS	SPIRES
DPS	MOLDS
RIMS	INFOL
DM-1	AIMS
ORBIT II	ALPS

3.2.3 Detailed Comparisons

Tables 4 through 7 contain the detailed comparisons of the bibliographic retrieval systems. The data contained are those items of interest for BCN planning with particular emphasis on those aspects of systems discussed in Section 2. The contents of the several tables are described below.

Table 4
The technical features of the systems are listed in Table 4. Included is information about the hardware, software, files, and I/O devices.

Table 5
Table 5 indicates which search strategies are available in each of the systems. The search strategies are divided into several categories: the retrieval terms on which retrieval may be performed; the items retrieved; the services provided; and whether or not using the system creates a keyword thesaurus. Explanation of the items retrieved follows:

On Original Search Term--(Self-explanatory.)

Synonyms--As provided for in a dictionary of synonyms.

Common Misspellings--As provided for in a table of misspellings.

By Statistical Association, etc.--According to some specified criteria where a term that is associated with the original search statement for some percentage of the searches would be retrieved.

*In some instances the differences may be considerable. For example, SPIRES is a retrieval system for physic literature whereas BALLOTS, which incorporates the retrieval system, also performs many library processing functions. For our purposes only the retrieval portion is considered.

Table 3. Overview of Selected Bibliographic Retrieval Systems

CHARACTERISTICS / SYSTEM	DEVELOPER	BRIEF DESCRIPTION AND SELECTED FEATURES	BIBLIOGRAPHIC APPLICATIONS	AVAILABILITY
BALLOTS Bibliographic Automation of Libraries on Time-sharing	Stanford Univ. Palo Alto, Calif.	Coordinate searching and retrieval of bibliographic data plus technical library functions. Computer system serves other university needs. Serials control planned. MARC tapes as basis for producing local files planned.	Initially physics citations (SPIRES) now includes general	Limited to Stanford U.
DATA CENTRAL	Data Corp.	Full text storage and retrieval system for various applications . Creates keyword dictionary as used.	First (COSATI Research Info) and Ohio Bar Case	Commercially available
DM-1 Data Management System	Auerbach, Inc.	Conversational file management system. Machine Independent. Operates on any of three computer systems. Users require computer programming orientation.	Various	Commercially available
DPS Document Processing System	Syracuse Univ.	Conversational bibliographic storage and retrieval system, primarily in a research environment. Primarily a remote batch system with limited on-line capability to be expanded in March or April, 1970	Psychological Abstracts, MARC	Limited to Syracuse U.
ELHILL	SDC	Conversational bibliographic storage and retrieval system based on ORBIT system (q.v.). Allows retrieval on various terms. Provides relevance feedback, tutorials, and other aids including on-line display of thesaurus and exploded search capability. Has rich command language and output selectivity. Used for NLM's AIM-TWX data base of over 100,000 citations.	NLM's AIM-TWX data base	Limited to NLM
INFORM-360	IBM	Store, disseminate, and retrieve text information used in conjunction with microfiche equipment. Microprocessing applications being investigated. Used with TEXTPAC for citation retrieval.	IBM documentation	Internal IBM system

Table 3. Overview of Selected Bibliographic Retrieval Systems

CHARACTERISTICS / SYSTEM	DEVELOPER	BRIEF DESCRIPTION AND SELECTED FEATRUES	BIBLIOGRAPHIC APPLICATIONS	AVAILABILITY
ORBIT II On-Line Bibliographic Information Transfer	SDC	Conversational bibliographic storage and retrieval system. Allows retrieval on various terms. Provides relevance feedback, tutorials, and other aids. Has rich command language and output selectivity. Technical Library Functions available Used as test vehicle for NLM planning.	NLM (Parkinson's Disease), MARC, ERIC and others	Commercially available
RECON	Lockheed Missile and Space Corp.	Conversational bibliographic storage and retrieval system.	NASA STAR citations and abstracts	Commercially available system (DIALOG version)
RIMS	Northwestern Univ.	Conversational bibliographic storage and retrieval system. Present system incorporates work on BIDAP, TRIAL, and INFOL. Later version to include statistical analysis capability.	Various test cases, technical literature	Limited to Northwestern Univ.
SPIRAL	Sandia Corp. Albuquerque, N.M.	Free flowing text storage and retrieval system that allows combined use of Boolean and relevance expressions. Current version is batch mode only, on-line version planned but no schedule set.	Test Data Base on Science and Technology	Internal Sandia System
SUNY BIO COM NET	State Universities at New York	Interactive bibliographic storage and retrieval system linking a number of libraries into a network. Also processes inter-library loan requests and performs many other library functions. This is essentially an on-line adaptation of the DPS system described above. It is included as a separate entry because of its biomedical orientation and because it is networked.	MEDLARS	Primarily SUNY and NLM

Terms in Same Structural Unit--Term A is retrieved only if it appears in the same (sentence, paragraph, etc.) as Term B.

Terms Adjacent, etc.--Term A is retrieved only if it is adjacent to (no more than X words from) Term B.

Terms with Similar Spelling, etc.--Provides ability to retrieve words that are likely to be close to the original term in the index. Usually presented as a listing of terms preceding and following the term of interest.

Other Associative Techniques--Includes a variety of techniques not otherwise classified, such as association based on similar index profiles.

The four items under the heading "Provides or Allows" denote the following:

Search Statement Linking--A search statement may be linked (by Boolean operators) to a previous search statement or earlier search statements may be linked to each other (although this is normally limited to the current session of system use).

Search Statement Order Independence--Search statements may be entered in any order and may be referred to or linked in any order.

Search Statement Truncation--Allows user to enter part of a search statement and truncate it using a special symbol. Especially useful for entering a long search term that is likely to be unique (such as a title).

Search Statement Qualification--Allows user to qualify a search statement when it is entered. For example, the search statement may be the author's name but the user may be interested in only those documents published since 1963. In this case, the date would be used as a qualifier and all earlier documents by the same author would be rejected.

The fact that a keyword thesaurus is created when the system is used is significant in that it provides the user with an additional search tool with which he can refer to, link, or otherwise search on terms in the new thesaurus.

Table 6

Table 6 shows a variety of additional on-line operating characteristics found in the systems. The items listed were selected on the basis that they would provide a reasonably complete picture of the several systems in terms of BCN planning interests. Thus, such features as security (which is available in some form in all systems) were not included, since they are of no concern

with respect to BCN bibliographic data.* A number of the terms used in Table 6 require further explanation. These are:

Accounting Techniques--Indicates that the system will provide information concerning its use (e.g., CPU time).

Arithmetic Operators--Indicates, without specifying which ones, that arithmetic (or statistical) operators are available for retrieval operations. Usually addition, subtraction, multiplication, division, and exponentiation are the minimal set and these may be applied to any retrieval categories that contain quantitative data. Note that an M in the column indicated those systems where arithmetic operators are available for file maintenance and not retrieval.

Boolean Operators--Includes "and", "or", and "not". An "M" indicates the Booleans are available for maintenance only.

Browse Mode of Operations--Allows users to browse through file. Generally involves display of several adjacent index terms so that user may find terms similar in spelling to the one in which he is interested.

Compose Printout Format--Allows user to specify a format in which he wants to have material output. Format is open ended and limited only by the capabilities of the remote terminal device.

Cueing or Prompting Signals--Presents a distinctive cue informing user that it is ready to accept a new input and may indicate type of input. Many systems also indicate when the computer is about to respond. Note that cueing signals assist in distinguishing between user input and system output.

Cue Level Select--Implies several levels of cueing signal which the user, based on his experience level, may select.

Exploded Search--Provides the ability to search on all terms subordinate to any index term. For use with data bases with hierarchically organized indexing vocabularies.

Graphics Capability--Indicates that the system has the ability to create and display graphics based upon file contents. Note the distinction between this and the next item.

*Security may become a problem if patient-record data are also handled by the system. If so, techniques to handle the problem are available.

Image Store and Retrieve--Indicates the ability to store and later retrieve graphics or image data. For example, a system that stores and retrieves microfiche would have such an ability. This ability does not include creating graphics from data stored in files.

Logical Comparators--(Also known as relational operators.) Includes such concepts as equal to, greater than, less than, not equal to, etc. Like arithmetic operators, in some systems logical comparators may be available for file maintenance and not retrieval; where this is the case, an M appears in the column.

Notification, Number of Items--Notifies the user of the number of items to be printed on-line prior to initiating printing or after printing a few items. Includes a requirement that the user respond positively that he indeed wants the number of items indicated (or some lessor number or none) before printing.

On-Line Thesaurus--Provides the ability to ascertain the relationship of a specified term to other terms in the data base. For use with data bases with structured relations among terms in the vocabulary.

Personal Terms Allowed (Rename Capability)--Allows use of personal language terms rather than requiring use of system-defined terms. For example, the user may substitute any term for some system command and then use that as the command name.

Remote Terminal Intercommunication--Allows users at remote terminals to communicate directly with each other in plain language via their remote terminal devices.

Qualify or Limit Printout--Allows the user to limit printout in terms of some value contained in the unit record. For example, the user might specify that he only wants records with publication dates later than 1965 or records published in English, and so on.

Search Hit Indicator--Indicates number of terms that matched search statement(s). Useful when search statements are linked by Boolean operators.

Select Items for Print--Allows user to select specified items from unit records for printing.

Self Destruct Protection--Protects the user from accidentally destroying his files. Usually implies system protects users from each other, i.e., they cannot damage each other's files. Also, they cannot damage the executive or operating system.

System Initiates Instructions--Initiates instructions telling the user what to do next. Similar to prompting or cueing but involves more in that prompting or cueing is usually limited to the retrieval functions, whereas for all system operations.

Trouble or Complaint Report--Provides standard means for entering trouble reports or gripes about system features.

Text Query--Provides the ability to query and retrieve from text material. This is distinct from Text Retrieval, where the text is retrieved regardless of whether a text query capability exists.

Tutorials on Demand--(Self-explanatory.)

Table 7

The error-control capabilities listed in Table 7 are mostly self-explanatory. Two items that may not be obvious are Variable Spelling Approximator and System Error Monitor and Record. The approximator operates when the search term cannot be found in the index file. It notifies the user and prints out similar terms. While this appears similar to common misspellings, it functins differently. The approximator finds similar words, i.e., neighboring terms, in the index. A misspelling dictionary may contain words that are very dissimilar or distant from the position of the search term in the index.

The system error monitor records errors. The record is then available to system operators or users. This can be especially valuable in providing feedback about errors that may occur frequently, and is particularly helpful during the early phases of a system development.

Table 4. Technical Features

CHARACTERISTICS / SYSTEM	CPU	OS OR SYST EXEC	NO. SIMULT USERS	CORE CAPACITY	USER/CORE ALLOCAT	MASS STORE CAPACITY (ON-LINE)	FILES						INTERACTIVE I/O DEVICES				CONVERSATIONAL LANGUAGE
							SIZE	MAX CAPACITY	ACCESS	UNIT RECORDS	INVERTED	LEVELS	TTY	2741 OR SIMIL	CCI OR SIMIL	OTHER	
BALLOTS	360/67	OS 360	--	350K	all	$27X10^6$ char.	--	VNL	R	Free Form; Map & Bit Masks	Y	1	N	Y	Y	--	UOL
DATA CENTRAL	360/40	DOS	50	65K partition	all	$10x10^9$ char.	65K fields	VNL	R	65K field of entry	Y	1	Y	Y	Y	--	UOL
DM-1	360/50 GE 635	OS 360 MVT HASP PCP	OS Dependent	--	65K bytes + 5K required by pgm.	1-2314 1-2311	--	VNL	R	free form	Y	any	(Function of Host Operating System)				UOL
DMS	365/50	OS 360	(primarily batch)	512K bytes	$24X10^6$ bytes	--	--	VNL	R	--	Y	1	Y	Y	N	--	DOC PROS
ELHILL	360/67	SDC TSS exec.	40	512K bytes	All	$432 X 10^6$ char.	$24X10^6$ English words	VNL	R	free form; var length & field	Y	16	Y	Y	Y	--	UOL
INFORM-360	360/50	OS 360	75	512K bytes	*	--	--	VNL	R	20bytes fixed 60 free	Y	1	N	Y	N	2260	UOL
ORBIT II	360/67	SDC TSS exec.	40	512K bytes	All	$432 X10^6$ char.	$24X10^6$ English words	VNL	R	Free Form; Var Length & field	Y	16	Y	Y	Y	--	UOL

*Supervisor assigns work area depending upon system load.

Table 4. Technical Features (cont.)

CHARACTERISTICS / SYSTEM	CPU	OS OR SYST EXEC	NO. SIMULT USERS	CORE CAPACITY	USER/CORE ALLOCAT	MASS STORE CAPACITY (ON-LINE)	FILES						INTERACTIVE I/O DEVICES				CONVERSATIONAL LANGUAGE
							SIZE	MAX CAPACITY	ACCESS	UNIT RECORDS	INVERTED	LEVELS	TTY	2741 OR SIMIL	CCI OR SIMIL	OTHER	
RECON	360/67 or any	OS 360	25	100K char (72K partition)	All	600 X10 6	--	VNL	R	Free Form Var. length & field	Y	User specified	Y	Y	Y	--	UOL
RIMS	CDC 6400	SCOPE 3.1.6	10	57K octal words	By user request	30-40 x 10 6 char.	--	VNL	R	full text	N	--	Y	N	N	CDC 200	UOL
SPIRAL	7090*	IBSYS	Batch	--	All	56K	--	{tape orient}	Serial	Full text	Y	--	N	N	N	N	VOL
	1108#	EXEC 2 EXEC 8	--	128K	All	64K	--	VNL	R				Y	N	N	N	VOL
SUNY BIO COM NET	360/40	OS 360	21	256K char.	165K	600 X10 6 char.	--	VNL	R	free form var. length & field	Y	1	N	Y	Y	--	DOC PROS

*Current version
#Planned

Table 5. Search Strategies

CHARACTERISTICS / SYSTEM		RAILOTS	DATA CENTRAL	IM-1	DRS	ERIL
RETRIEVAL TERMS	Author	Y	Y	Y	Y	Y
	Title	Y	Y	Y	Y	Y
	Keywords/Index Terms	Y	Y	Y	Y	Y
	Reference Number	Y	Y	Y	Y	Y
	Other	--	any term	--	--	Vari-ous
RETRIEVES	On Original Search Term	Y	Y	Y	Y	Y
	Synonyms/Root Words	N	N	N	Y	N
	Common Misspellings	N	N	N	Y	N
	By Statistical Association With Original	N	N	N	N	N
	Terms In Same Structural Unit (sentence, para, etc.)	N	Y	N	N	N
	Terms Adjacent or Specified No. from Original Term	N	Y	N	N	N
	Terms With Similar Spelling (Adjacent in Index)	N	N	N	Y	Y
	By Other Associative Techniques.	N	N	N	N	N
PROVIDES OR ALLOWS	Search Statement Linking	Y	Y	(can be included in applications programs)	Y	Y
	Search Statement Order Independence	Y	Y		Y	Y
	Search Statement Truncation	Y	Y		Y	Y
	Search Statement Qualification	Y	Y		Y	Y
Creates Keyword Thesaurus		N	Y	N	N	N

ABBREVIATIONS
Y--Yes
N--No

Table 5. Search Strategies

CHARACTERISTICS / SYSTEM	RETRIEVAL TERMS: Author	Title	Keywords/Index Terms	Reference Number	Other	RETRIEVAL ACHIEVES: On Original Search Term	Synonyms/Root Words	Common Misspellings	By Statistical Association With Original	Terms In Same Structural Unit (sentence, para, etc.)	Terms Adjacent or Specified No. from Original Term	Terms With Similar Spelling (Adjacent in Index)	By Other Associative Techniques.	PROVIDES OR ALLOWS: Search Statement Linking	Search Statement Order Independence	Search Statement Truncation	Search Statement Qualification	Creates Keyword Thesaurus
INFORM-360	Y	Y	Y	Y	any term	Y	Y	N	N	N	N	N	N*	Y	Y	Y	Y	Y
ORBIT II	Y	Y	Y	Y	Various	Y	N	N	N	N	N	Y	N	Y	Y	Y	Y	N
RECON	Y	Y	Y	Y	--	Y	N	N	N	N	N	Y	N	Y	Y	Y	Y	N
RIDS	Y	Y	Y	Y	numeric data	Y	N	N	N	N	N	Y	N	Y	Y	Y	Y	N
SPIRAL	L	L	Y	Y	text	Y	L	N	N	N	N	N	N	Y	Y	Y	Y	N
SUNY BIO COM NET	Y	Y	Y	Y	--	Y	Y	Y	N	N	N	Y	N	N	N	N	N	Y

ABBREVIATIONS
Y--Yes
N--No
L--Limited

Notes
(1) Phrases may be scrambled
(2) Model paragraph involves keyword scoring

* Uses pattern matching. Each entry word essentially "or" linked to all others.

Table 6. On-Line Operating Characteristics

CHARACTERISTICS / SYSTEM	BALLOTS	DATA CENTRAL	DM-1	DPS	ELHILL
Abstract Retrieval	Y	Y	N	N	A
Accounting Techniques	Y	Y	N	N	Y
Arithmetic Operators	N	Y	Y	N	Y
Boolean Operators	Y	Y	Y	Y	Y
Browse Mode of Operations	Y	Y	Y	Y	Y
Cueing/Prompting Signals	Y	Y	N	Y	Y
Cue Level Select	N	N	N	N	N
Compose Printout Format	Y	Y	Y	Y	N
Direct Micro Image Read	N	N	N	N	N
Direct Micro Image Write	N	N	N	N	N
Direct Optical Scan Read	N	N	N	N	N
Direct Optical Scan Write	N	N	N	N	N
Exploded Search	N	N	N	N	Y
Full-Text Retrieval	N	Y	N	N	N
Graphics Capability	N	N	N	N	A
Image Store & Retrieve	N	N	N	N	N
Logical Comparators	Y	Y	Y	Y	Y
Notification, No. of Items	Y	Y	Y	Y	Y
On-Line File Maintain/Update	N	N	N	Y	N
On-Line Indexing	N	N	Y	Y	N
On-Line Thesaurus	N	Y	N	N	Y
Past Query File	Y	Y	N	--	Y
Personal Terms Allowed	N	N	N	--	Y
Qualify or Limit Printout	Y	Y	Y	Y	Y
Search Hit Indicator	Y	Y	Y	Y	Y
Select Items for Printout	N	Y	Y	--	Y
Self-Destruct Protection	N	Y	Y	N	N
SDI or Other Dissemination	N	Y	Y	N	N
Status Display on Demand	Y	N	N	N	Y
System Initiates Instr's	Y	N	N	Y	Y
Text Query	N	Y	N	N	Y
Trouble or Complaint Report	Y	N	Y	--	A
Tutorials on Demand	N	N	Y	--	Y

ABBREVIATIONS

Y--Yes
N--No
P--Planned
U--Unknown

NA--Not Applicable
F--File Maintenance Only
A--Available

Table 6. On-Line Operating Characteristics (Cont'd)

CHARACTERISTICS / SYSTEM	INFORM-360	ORBIT II	RECON	RDMS	SPIRAL	SUNY BIO COM NET
Abstract Retrieval	Y	A	Y	Y	Y	N
Accounting Techniques	N	Y	Y	Y	N	N
Arithmetic Operators	Y	Y	Y	Y	N	N
Boolean Operators	Y	Y	Y	Y	Y	Y
Browse Mode of Operations	Y	Y	Y	Y	N	Y
Cueing/Prompting Signals	Y	Y	N	Y	N	Y
Cue Level Select	Y	Y	N	N	N	N
Compose Printout Format	N	N	Y	N	N	Y
Direct Micro Image Read	Y	N	N	N	N	N
Direct Micro Image Write	Y	N	N	N	N	N
Direct Optical Scan Read	N	N	N	N	N	N
Direct Optical Scan Write	N	N	N	N	N	N
Exploded Search	N	N	N	N	N	N
Full-Text Retrieval	Y	N	N	N	Y	N
Graphics Capability	N	A	N	N	N	N
Image Store & Retrieve	N	N	N	N	N	N
Logical Comparators	Y	Y	N	Y	N	Y
Notification, No. of Items	Y	N	Y	Y	N	Y
On-Line File Maintain/Update	N	N	N	N	N	Y
On-Line Indexing	N	Y	N	N	N	Y
On-Line Thesaurus	N	N	N	N	N	N
Past Query File	Y	Y	Y	N	N	-
Personal Terms Allowed	N	Y	N	N	N	-
Qualify or Limit Printout	Y	Y	Y	Y	Y	Y
Search Hit Indicator	Y	Y	Y	Y	Y	Y
Select Items for Printout	Y	Y	Y	Y	Y	Y
Self-Destruct Protection	Y	Y	Y	Y	Y	-
SDI or Other Dissemination	N	N	Y	N	N	N
Status Display on Demand	-	Y	-	-	N	N
System Initiates Instr's	Y	Y	-	N	N	Y
Text Query	Y	Y	N	Y	Y	N
Trouble or Complaint Report	Y	A	Y	N	N	-
Tutorials on Demand	Y	Y	N	Y	N	-

ABBREVIATIONS

Y--Yes
N--No
P--Planned
U--Unknown
NA--Not Applicable
F--File Maintenance Only
A--Available

Table 7. Error-Control Capabilities

CHARACTERISTICS / SYSTEM	SYSTEM PERFORMS THE FOLLOWING (YES OR NO)							
	Displays Entry Preparation	Edit Entry Capability	Explanatory Error Message	Detailed Error Messages on Demand	No-loss Recovery*	Variable Spelling Approximator	Cancel or Abort Active Procedure	System Error Monitor and Record
BALLOTS	Y	Y	Y	Y	Y	N	Y	--
DATA CENTRAL	Y	Y	Y	Y	Y	Y	Y	Y
DM-1	Y	Y	Y	Y	Y	N	--	A
DPS	Y	Y	--	--	Y	Y	N	Y
ELHILL	Y	Y	Y	Y	Y	Y	Y depends upon terminal device	Y
INFORM-360	Y	Y	Y	Y	Y	Y	Y	Y
ORBIT II	Y	Y	Y	Y	Y	Y	Y depends upon terminal device	Y

*Following system cut-off, communications failure, or input error.

Table 7. Error-Control Capabilities (Cont'd)

CHARACTERISTICS / SYSTEM	SYSTEM PERFORMS THE FOLLOWING (YES OR NO)						
	Displays Entry Preparation	Edit Entry Capability	Explanatory Error Message	Detailed Error Messages on Demand	No-loss Recovery*	Variable Spelling Approximator	Cancel or Abort Active Procedure	System Error Monitor and Record
RECON	Y	Y	Y	Y	Y	Y	N	Y
RIMS	Y	Y	Y	Y	Y	N	Y	N
SPIRAL	N	N	N	N	Y	N	Y	N
SUNY BIO COM NET	Y	Y	--	--	Y	Y	N	Y

*Following system cut-off, communications failure, or input error.

3.3 COMMENTS ON THE COMPARISONS

The explicit structure that was imposed in the several tables in this section is based upon our conception of the variables important for BCN planning. The result is that the systems are compared along dimensions that may not be applicable for more general evaluations. This precludes overall comparative evaluations of the systems and no overall system evaluations were intended or performed.

In interpreting the comparisons in Tables 3 through 7 an important consideration is that the data are often limited to noting the presence or absence of certain system features. No comparisons are made of how well those features operate when they are present, or the efficiency of their underlying programs. For example, five systems provide an on-line tutorial feature, and this is noted; the nature of the tutorials, however, varies from system to system. For BCN purposes, one of these tutorials may well be better than the others, but this generally cannot be determined from document sources and is not indicated in the tables.

The general-purpose systems reviewed operate on a number of different computers, have diverse applications packages associated with them, and use various programming languages. They nevertheless display considerable similarity. This is especially true of their user-oriented features. They all tend to use the same types of remote terminals. Most have no limit on file size except that which is imposed by physical limits of the hardware. Most appear reasonably efficient with respect to file manipulation and report generation. Finally, while almost all can be used for bibliographic data storage and retrieval if special files and appropriate indexes are created, they are inefficient for this purpose.

The bibliographic retrieval systems are in many superficial respects similar to the general-purpose systems: they operate on the same types of computers, use the same types of terminals, and have many similar operating characteristics. They differ primarily in that their files are specially structured for handling bibliographic records and they facilitate using search strategies for retrospective searches. With respect to their overall or general features, the eleven bibliographic retrieval systems reviewed differ in a few important respects. One is the fact that some are capable of full-text storage and retrieval (Data Central, INFORM-360, and SPIRAL), whereas the others are not, although many of them could be modified to include full-text storage and retrieval. But their designers have opted not to do so. This is a deliberate decision based upon different philosophies about search and retrieval. Seven of the eleven systems (BALLOTS, DPS, ELHILL, INFORM-360, RIMS, SPIRAL, and SUNY) are available on only a limited or restricted basis; the other four (DATA CENTRAL, IM-1, ORBIT II, and RECON) are commercially available. All but

three (DPS, INFORM-360, and SPIRAL) operate in a time-shared environment, but time-sharing is planned or potentially available for those three.

All of the bibliographic retrieval systems reviewed provide sufficient core capacity and on-line storage to allow complex search and retrieval of very large data bases. The capacities shown in Table 4 reflect current equipment configurations; the system capacities can be significantly increased by simply adding new equipment. The limit is thus more economic than technical.

The search strategies incorporated into the eleven systems are similar in that they all provide retrieval on author, title, index terms, and reference numbers; only ORBIT II and ELHILL provide retrieval on a wide variety of additional categories. The three full-text systems, as the name implies, allow retrieval on essentially any terms. The systems all have such special features as search-statement linking, and most allow search-statement truncation. Few have any ability to retrieve by any technique other than the original search terms.*

For BCN planning, there are interesting differences in the operating characteristics or other technical features of the systems. Of particular interest is a spelling approximator (i.e., misspelled search terms can be meaningfully interpreted by the system) found in seven systems (DATA CENTRAL, DPS, ELHILL, INFORM-360, ORBIT II, RECON, and SUNY); and tutorials on demand, found in five systems (DM-1, ELHILL, INFORM-360, ORBIT II, and RIMS). (BALLOTS) performs additional library technical functions, such as cataloging, circulation, or serials control, in addition to the storage and retrieval functions. Such functions can also be performed in conjunction with ORBIT II by a companion program known as ALPS which is not described in this report. Three systems (DATA CENTRAL, DM-1, and RECON) provide an SDI capability, although this would be readily available in all systems with minor modification. One system (INFORM-360) is designed specifically to produce microforms. The error control procedures of all the systems are essentially similar. Two systems (DATA CENTRAL and ELHILL) present on-line thesaurus data. Only ELHILL provides a hierarchically expanded search capability.

Two reasonably consistent "models" of bibliographic data retrieval systems emerge from the comparisons. The major difference between them is their file structure--one is oriented toward bibliographic citations, the other toward full text. Beyond that, there are differences in the purposes or applications

*Work on some experimental systems have been in the forefront of associative and statistical techniques. A noteworthy example is the work being performed by Salton in Project SMART at Cornell. The techniques being developed there may be expected to become available at the time of the BCN implementation. Their cost/effectiveness implications for the BCN, however, is not yet apparent.

for which the systems were designed, differences in their availability, and the differences in on-line operating characteristics just described.

4. IMPLICATIONS FOR THE BCN

The implications for the BCN contained in this section are based upon the review, analysis, and comparison of the systems. They support the basic BCN concepts described in the NLM/BCN plan. They are presented in terms of considerations for the system design, BCN users, and network planning.

4.1 IMPLICATIONS FOR SYSTEM DESIGN

A major consideration in the design of the BCN in whether or not to provide a full-text retrieval capability. The analysis in Section 2.3 indicated that the cost of full-text retrieval in terms of file size and structure (and ultimately in money) tends to be prohibitive for very large data bases. In addition, many of the advantages of full-text retrieval can be obtained by providing adequate on-line interactive search capability. Finally, those who may be expected to use the BCN are not likely to want or to use a full-text capability. Therefore, we strongly recommend that NLM stay with the present design concept and not consider full-text capability. This recommendation has a corollary recommendation for providing certain interactive capabilities for users, which are described in Section 4.2.

Another important consideration will be the choice between linking several existing systems and developing a completely new system. The former approach would entail the considerable task of building interfaces among the linked systems. Not only must all system elements be made compatible with each other, but modifications would be required to the executive or operating systems of the linked computers in the systems. Moreover, the difference in search strategies among the systems could create problems of training transferability when operating at different sites. In view of all this, we suggest that a unified network concept be designed and developed. Existing systems may sometimes be "plugged into" the network; where this is done, it should be on an interim basis, and degraded performance should be expected and allowed.

Standards and specifications for inputs, thesauri, unit records, and procedures need to be specified as part of network design and development. The experience with NLM applications found in the ORBIT II, ELHILL, and SUNY systems should provide one important basis for the standards and specifications. Also, file structure optimization is necessarily related to specifications for inputs and thesauri. This is more fully discussed in Section 2.3 on pages 7 through 13.

4.2 IMPLICATIONS FOR THE BCN USER

The present BCN design concept is oriented toward having librarians or other information specialists perform the remote terminal transactions. While this approach is adequate for many users, it probably fails to exploit the potential that accrues by using the system in a direct interactive mode. As stated on page 16 of this report, "...on-line interactive users who receive immediate feedback from the system are likely to receive many cues which will lead them in directions not originally anticipated; this results in an adaptive or heuristic quality that further enhances searching. Users adjust their search strategy in terms of immediate relevance judgments." We believe that, for this reason, some users will want to take advantage of the direct interactive mode. Therefore, more emphasis should be placed on that capability in the system design. Indeed, the on-line interactive capability may, in the long run, result in better retrieval at less cost than elegant retrieval strategies that lack the interactive aspects.

The user procedures and language should be consistent throughout the system. This is another reason for not linking existing systems (except those that may allow the use of a standard language and set of procedures). In addition, the network design should incorporate the twelve items contained in Section 2.4, User Considerations, on pages 17 and 18.

4.3 IMPLICATIONS FOR NETWORK PLANNING

To the extent that the network incorporates elements not specifically designed as part of the BCN, there will be an increased need for attention to standardization and convertability. That is, where different computers or programs must be linked together, additional costs are increased. Either standards must be imposed, conversion routines or algorithms written, or interfacing hardware installed in order to provide compatibility in such things as word size, data rates, or data formats. Thus, this review provides support to the approach to compatibility taken in the NLM/BCN Development Plan.

An important problem for the BCN is that of the potentially high cost of file access and file duplication. Each local center could have a duplicate of the entire file. On the other hand, there could be one central file accessible by remote terminals at each center. In the first case, the main costs are those associated with storage capacity, file size, and the cost of multiple file updating; in the second case, the main costs are those associated with communications and interface requirements. The system could operate in a mixed mode, with some part of the file duplicated at each center and other

parts maintained only centrally. This would facilitate efficient use of storage space by allowing data bases to be tailored to the user set at a location. This approach is probably the most efficient, and is, therefore, recommended for the network design.

5. BIBLIOGRAPHY

1. ATHERTON, PAULINE, and WYMAN, JOHN. "Searching MARC Project Tapes Using IBM/Document Processing System." Proceedings of the American Society for Information Science. Volume 6. Cooperating Information Societies. 32nd Annual Meeting, San Francisco, California, October 1-4, 1969. Greenwood Publishing Corporation, Conn. p. 83-88.

2. AUERBACH INFORMATION, INC. Auerbach Time Sharing Reports. Philadelphia, Pa., 1969.

3. BECKER, JOSEPH & HAYES, ROBERT M. Information Storage and Retrieval: Tools, Elements, Theories. John Wiley & Sons, New York, 1963.

4. BECKER, JOSEPH and OLSEN, WALLACE C. "Information Networks." Annual Review of Information Science and Technology. Volume 3, 1968. Encyclopaedia Britannica. Chicago. p. 289-327.

5. BORKO, H. The Analysis and Design of Information Systems. SDC SP-2655. System Development Corporation, Santa Monica, California, 10 November 1966.

6. CAPONIO, JOSEPH F. "NINDS Neurological Information Network." Proceedings of the American Society for Information Science. Volume 6. Cooperating Information Societies. 32nd Annual Meeting, San Francisco, California, October 1-4, 1969. Greenwood Publishing Corporation, Conn. p. 447-452.

7. CODASYL SYSTEMS COMMITTEE TECHNICAL REPORT. A Survey of Generalized Data Base Management Systems. Association For Computing Machinery, New York, May 1969.

8. CORNELL UNIVERSITY. DEPARTMENT OF COMPUTER SCIENCE. Information Storage and Retrieval. Scientific Report. No. IRS-14 to The National Science Foundation. Reports on Analysis, Search and Iterative Retrieval. Gerard Salton, Project Director. Ithaca, New York, October 1968.

9. CORNELL UNIVERSITY. DEPARTMENT OF COMPUTER SCIENCE. Information Storage and Retrieval. Scientific Report No. ISR-15 to The National Science Foundation. Gerard Salton, Project Director. Ithaca, New York, January 1969.

10. DATA CORPORATION. An Automated, General Purpose Information Storage and Retrieval System. Data Corporation, Arlington, Va., 1968.

11. DOWKONT, ANTHONY, J., MORRIS, WILLIAM A., and BUETTEL, T. DWIGHT. A Methodology for Comparison of Generalized Data Management Systems: PEGS (Parametric Evaluation of Generalized Systems), March 1967.

12. FANO, R. M. The MAC System: A Progress Report. Massachusetts Institute of Technology Project MAC. MAC-TR-12. October 9, 1964.

13. Fry, J. P. ET. AL. Data Management System Survey. MTP-329. The MITRE Corp., Washington, D.C., January 1969.

14. GIERING, RICHARD H. Analysis of Existing and Proposed Data Handling Systems. Technical Note. DTN-67-9. Data Corporation, Arlington, Va., 23 October 1967.

15. GRUENBERGER, FRED (Ed.). Critical Factors in Data Management. Prentice-Hall, Englewood Cliffs, N.J., 1969.

16. HEAD, ROBERT V. "Management Information Systems: A Critical Appraisal." Datamation May, 1967. p. 22-27.

17. HENDERSON, MADELINE M. Evaluation of Information Systems: A Selected Bibliography With Informative Abstracts, Technical Note 297. U. S. Department of Commerce, National Bureau of Standards, Washington, D.C., December 1967.

18. KATTER, ROBERT V., and BLANKENSHIP, DONALD A. On-Line Interfaces for Document Information Systems; Considerations for the Biomedical Communications Network. SDC TM-(L)-4320. System Development Corporation, Santa Monica, California. 3 June 1969.

19. KATTER, ROBERT V. "Design and Evaluation of Information Systems." Annual Review of Information Science and Technology. Volume 4, 1969. p. 31-70. Encyclopaedia Britannica, Inc. Chicago.

20. KING, DONALD W. "Design and Evaluation of Information Systems." Annual Review of Information Science and Technology. Volume 3, 1968. Encyclopaedia Britannica, Inc., Chicago. p. 61-103.

21. LEITER, JOSEPH, and GULL, C. D. "The MEDLARS System in 1968." Proceedings of the American Society for Information Science. Annual Meeting, Volume 5, Information Transfer. Columbus, Ohio. October 20-24, 1968. Greenwood Publishing Corporation, N.Y. p. 255-262.

22. LEVY, RICHARD P. and CAMMARN, MAXINE R. "Information Systems Applications in Medicine." Annual Review of Information Science and Technology. Volume 3, 1968. Encyclopaedia Britannica, Chicago. p. 397-428.

23. MASSACHUSETTS INSTITUTE OF TECHNOLOGY. Project Intrex, Semiannual Activity Report, 15 March 1969 to 15 September 1969, PR-8, Cambridge, Mass., 15 September 1969.

24. MERROW, CLYDE. On-Line File Manipulation Techniques. Technical Report No. RADC-TR-67-600. Rome Air Development Center, USAF, February 1968.

25. NATIONAL LIBRARY OF MEDICINE. Evaluation of the MEDLARS Demand Search Service. Washington, D.C., January 1968.

26. NATIONAL LIBRARY OF MEDICINE. The Biomedical Communications Network: Technical Development Plan. NLM, Bethesda, Md., June 1968.

27. NATIONAL SCIENCE FOUNDATION. OFFICE OF SCIENCE INFORMATION SERVICE. Current Research and Development in Scientific Documentation, No. 15. NSF-69-8. 1969.

28. OVERHAGE, CARL F. J. "Information Networks." Annual Review of Information Science and Technology. Volume 4, 1969. p. 339-377. Encyclopaedia Britannica, Inc. Chicago.

29. PIETRZYK, ALFRED, LAMBERTS, FRANCES, and FREEMAN, ROBERT R. File-Management Techniques and Systems with Applications to Information Retrieval - A Selective Bibliography. Center for Applied Linguistics, Washington, D.C. June 1968 (PB 178 792).

30. REILLY, KEVIN D. Evaluation of Generalized File Management Systems. 15 December 1967. p. 46. Institute of Library Research, University of California at Los Angeles.

31. SABLE, JEROME D., et. al. Data Management System Study. N68-26292. Auerbach Corporation, Philadelphia, Pa., 1 April 1968.

32. SALTON, GERARD. Information Storage and Retrieval. Cornell University, Ithaca, N.Y., June 1967, (PB 176 536).

33. SALTON, G. "The Evaluation of Computer-Based Information Retrieval Systems." Proceedings of the FID Congress, 1965. p. 125-133.

34. SARACEVIC, TEFKO, et al. An Inquiry into Testing of Information Retrieval Systems. Part 1. Objectives, Methodology, Design, and Controls. Case Western Reserve University, Ohio. 1968, (PB 179 290).

35. SCHECTER, GEORGE, Editor. Information Retrieval. A Critical View. Based on Third Annual Colloquium on Information Retrieval. May 12-13, 1966, Phila. Thompson Book Company, Washington, D.C., 1967.

36. SCROGGINS, J. L. MIZOUE, J. K., and GLASS, R. D. National Library of Medicine On-Line Retrieval of Bibliographic Information Text. NLM/ORBIT. SDC TM-WD-731/000/00. System Development Corporation, Falls Church, Va., 8 November 1968.

37. SENKO, MICHAEL E. "File Organization and Management Information Systems." Annual Review of Information Science and Technology. Volume 4, 1969. p. 111-143. Encyclopaedia Britannica, Inc., Chicago.

38. SHOFFNER, RALPH M. "Organization, Maintenance and Search of Machine Files." Annual Review of Information Science and Technology. Volume 3, 1968. Encyclopaedia Britannica, Inc., Chicago. p. 137-167.

39. SMITH, STEPHEN F. & SHOFFNER, RALPH M. A Comparative Study of Mechanized Search Languages. Institute of Library Research, University of California, January 1969.

40. SPARKS, DAVID E., CHODROW, MARK M., and WALSH, GAIL M. A Methodology for the Analysis of Information Systems. Final Report. Information Dynamics Corporation, Wakefield, Mass., May 1965. (PB 168 264)

41. STANFORD RESEARCH INSTITUTE. Requirements, Criteria, and Measures of Performance of Information Storage and Retrieval Systems. Final Report. SRI Project No. 3741. December 1961.

42. SYSTEM DEVELOPMENT CORPORATION. Technology and Libraries. SDC TM-3732. SDC, Santa Monica, Calif., November 1967.

43. SWETS, JOHN A. Effectiveness of Information Retrieval Methods. June 1967. Bolt Bernaek and Neuman Inc., Mass. (AD 656 340).

44. TIME-SHARING ENTERPRISES, INC. Time-Sharing Industry Directory. King of Prussia, Pa. 1969.

45. WARREN, PETER A., VINKEN, PIERRE J., and VAN DER WALLE, FRANS. "Design and Operation of an Advanced Computer System for the Storage, Retrieval and Dissemination of the World's Biomedical Information." Proceedings of the American Society for Information Science, Volume 6. Cooperating Information Societies. 32nd Annual Meeting, San Francisco, California, October 1-4, 1969. Greenwood Publishing Corporation, Conn. p. 423-429.

46. WELCH, NOREEN O. A Survey of Five On-Line Retrieval Systems. MTP-322. The Mitre Corporation, Bedford, Mass., August, 1968.

47. WESTAT RESEARCH, INC. Evaluation of Retrieval Systems: Literature Perspective, Measurement, Technical Papers. Bethesda, Md., December 1968. (PB 182 710).

48. ZIEHE, T. W. Data Management-A Comparison of System Features. Report of Project 007 001 01. Document No. Tracor 67-904-U. Austin, Texas, TRACOR, 1967. (AD 661 861).